DEPARTMENT OF ELECTRICAL ENGINEERING
SIGNAL PROCESSING LABORATORY
Professor M. Kunt - Director

EPFL–DE–LTS CH–1015 LAUSANNE SWITZERLAND
Telephone: +41-21-693-2626 Secretariat: +41-21-693-2601
Telefax: +41-21-693-7600
Email: kunt@epfl.ch

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Joint Source/FEC Rate Selection for Quality-Optimal MPEG-2 Video Delivery

Pascal Frossard and Olivier Verscheure

Internal Report LTS 99.04
Lausanne, October, 1999

Joint Source/FEC Rate Selection for Quality-Optimal MPEG-2 Video Delivery

Pascal Frossard and Olivier Verscheure[1]

Signal Processing Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland

pascal.frossard@epfl.ch

Contents

[1]O. Verscheure is with the IBM T.J. Watson Research Center, New York, USA. E-mail: ov1@us.ibm.com

Abstract

This paper deals with the optimal allocation of MPEG-2 encoding and media-independent Forward Error Correction (FEC) rates under a total given bandwidth. The optimality is defined in terms of minimum perceptual end-to-end distortion given a set of video and network parameters. We first derive the set of equations leading to the residual loss process parameters. That is, the packet loss ratio and the average burst length after FEC decoding. We then show that the perceptual source distortion decreases exponentially with the increasing MPEG-2 source rate. We also demonstrate that the perceptual distortion due to data loss is directly proportional to the number of lost macroblocks, and therefore decreases with the amount of channel protection. Finally, we derive the global set of equations that lead to the optimal dynamic rate allocation. The optimal distribution is shown to outperform classical FEC scheme, thanks to its adaptivity to the scene complexity, the available bandwidth and the network performances.

I. Introduction

Streamed digital video has already begun its penetration in the market. Transmitting video in digital form is the direct result of the benefits offered by digital compression. The purpose of compression is data rate reduction, which results in lower transmission costs. Although many compression schemes have been studied, a standard was necessary for widespread communications. For digital video, attention is now being focused on the MPEG-2 standard. MPEG-2 aims at diverse applications such as television broadcast over satellite, cable and other broadcast channels (DVB), and digital storage media (DVD).

In general, the distortion the end-user perceives results from compression artifacts, packet losses, delays and delay jitters. All lossy compression schemes distort and delay the signal. Degradation mainly comes from the quantization, which is the only irreversible process in a coding scheme. Moreover, delays and packet losses are inevitable during transfers across today's networks. The delay is generally due to propagation and queuing. Information loss is mainly caused by multiplexing overloads of high magnitude and duration that lead to buffer overflow in the nodes. Data loss is particularly annoying in video streaming applications due to the predictive structure of MPEG-2 compression, as described in the next section.

Interactive video delivery can significantly be improved by providing sender-side mechanisms [1]. These include (i) structuring techniques [2–4] and scalable coding [5, 6] to reduce data loss sensitivity, and (ii) forward error correction (FEC) mechanisms to lower the probability of loss at the application layer. FEC means that redundancy is added to the data so that the receiver can recover from losses or errors without any further intervention from the sender. This work explores the media-independent FEC scheme [7]. That is, k video packets are protected by $n-k$ FEC packets within a total of n, regardless of the underlying video sensitivity [8].

Clearly, under a given channel rate, the addition of FEC packets reduces the available rate for source coding. Rates allocated to source coding and FEC results thus from a trade-off whose optimisation is the topic of this paper. Also, this optimal distribution has to be dynamically adjusted according to varying video and network parameters.

The problem is therefore stated in the following manner. Let $R(t)$ denote the channel rate available for transmission at time t for a given time slot. Let $R_S(t)$ and $R_{FEC}(t)$ further denote the MPEG-2 source rate and the rate of FEC packets. The problem is then to *find* the optimal $R_S(t)$ and $R_{FEC}(t)$ at time t that *minimize* the end-to-end video distortion[2] *under* the constraint $R_S(t) + R_{FEC}(t) \leq R(t)$ and a given set of video and network parameters.

The paper is organized as follows: Section II briefly describes the transmission of

[2]We call "end-to-end video distortion" the distortion as perceived by the end-user.

MPEG-2 video streams over packet networks and the sensitivity to data loss. Section III analyzes the FEC efficiency. The residual video loss patterns after FEC reconstruction are computed in the case of a Gilbert-model loss process. The source perceptual distortion-rate function is empirically computed in Section IV. Section V then analyzes the error propagation and derives the degradation due to loss from the video loss patterns. The optimal rate distribution is the topic of Section VI. Section VII shows how our algorithm behaves in different conditions, and compares the end-to-end distortion to classical FEC schemes. Finally, concluding remarks are given in Section VIII.

II. MPEG-2 over Packet Networks

In this section, we first briefly present the MPEG-2 encoding algorithm. We then explain why data loss is particularly annoying in MPEG-2 streaming applications.

A. MPEG-2 Backgrounder

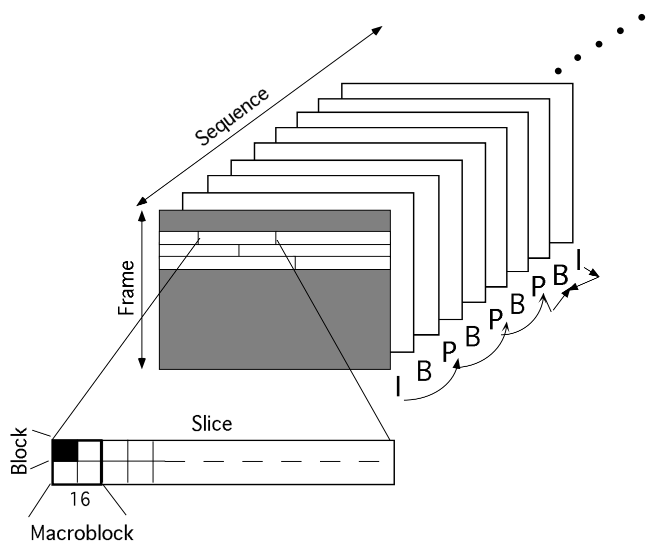


Fig. 1. MPEG-2 video structure.

An MPEG-2 video stream is hierarchically structured as illustrated in Fig. 1. The stream consists of a sequence composed of several frames. The MPEG-2 video standard defines three different types of frames: intra-coded (I-), predicted (P-) and bidirectional (B-) frames. The use of these three frame types allows MPEG-2 to be robust as I-frames provide error propagation reset points and efficient as B- and P-frames allow a good overall compression ratio. Each frame is composed of slices which are series of macroblocks. Each macroblock (16×16 pixels) contains 4 blocks (8×8 pixels) of luminance and 2, 4 or 8 blocks of chrominance depending on the chroma format. Motion estimation is performed on macroblocks while the DCT [3] is calculated on blocks. The resulting DCT coefficients are quantized and variable length coded. The quantizer results from the multiplication of a Quantizer Scale, MQUANT, and the corresponding element of a Quantizer Matrix. In general, the higher the MQUANT value, the lower the bit rate but also the lower the quality (well-known from the rate-distorsion theory).

Before being transmitted, a video stream goes through the MPEG-2 Transport Stream (TS) layer. Basically, the stream is first segmented into variable-length Packetized Elementary Stream (PES) packets and then subdivided into fixed-length TS packets. These packets are then encapsulated in RTP (Real-Time Protocol) packet for transmission [9,10]. It is worth noting that a uncompressed header (i.e., syntactic information) is inserted before each of the following information elements: PES, TS, sequence, picture, and slice. In general, when a header is damaged, the underlying information is lost.

B. *MPEG-2 Sensitivity to Data Loss*

Figure 2 illustrates how network losses map onto visual information losses in different types of pictures. Data loss spreads within a single picture up to the next resynchronization point (e.g., picture or slice headers) mainly due to the use of differential coding, run-length coding and variable length coding. This is referred to as spatial propagation and may damage any type of picture. When loss occurs in a reference picture (intra-coded or predictive frame), the damaged macroblocks will affect the non intra-coded macroblocks in subsequent frame(s), which reference the errored macroblocks. This is due to inter-frame predictions and known as temporal propagation.

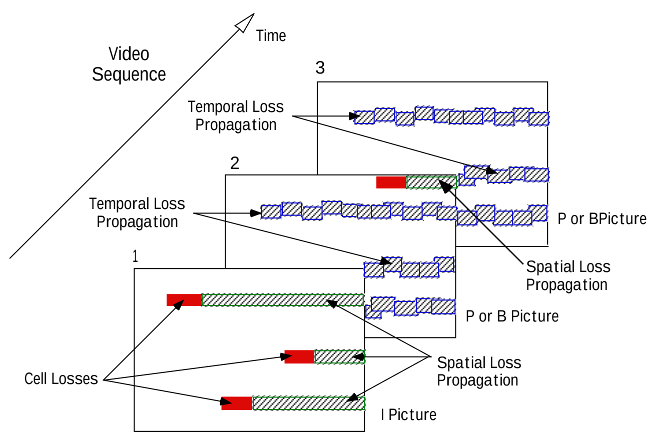


Fig. 2. Data loss propagation in MPEG-2 video streams.

However, the error visibility may be dramatically reduced by means of error concealment techniques [11]. These error concealment algorithms include, for example, spatial

[3]DCT stands for Discrete Cosine Transform

interpolation, temporal interpolation and early resynchronization techniques [12]. The MPEG-2 standard proposes an elementary error concealment algorithm based on motion compensation. Basically, it estimates the motion vectors for the lost macroblock by using the motion vectors of neighbouring macroblocks in the affected picture (provided that these have not also been lost). This improves the concealment in moving picture areas. However, there is an obvious problem with errors in macroblocks whose neighbouring macroblocks are intra-coded, because there are ordinarily no motion vectors associated with them. To circumvent this problem, the encoding process can be extended to include motion vectors for intra macroblocks[4].

In general, error concealment techniques may, in general, efficiently decrease the sensitivity to data loss. However, none of these techniques is perfect. Data loss may still involve annoying degradation in the decoded video.

III. Loss Process Parameters after FEC Recovery

Forward Error Correction (FEC) techniques are the preferred error-control scheme for multicast or interactive applications. In this paper a very simple media-independent Forward Error Correction (FEC) mechanism is used. Due to the low bit error rates associated with the modern communication media, the assumption is made that decoding is mainly impeded by packet loss. Either the packet is present and correct or it is lost. These losses are mainly caused by network congestion and the resultant buffer overflow and queuing delay. In this case, packet-level FEC schemes [13–16] provide an efficient way to fight against losses, although the perfect recovery cannot be guaranteed. The details of the FEC algorithms are outside the scope of this paper but can be found in [17]. Recall however that common FEC schemes based on Reed-Solomon codes or X-OR functions can generally correct as many losses as the number of redundancy packets.

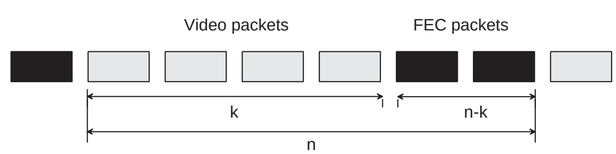


Fig. 3. Media-independent FEC scheme.

Assume every block of k video packets are protected by $(n-k)$ FEC packets as represented in Figure 3 in a n RTP packets stream. This is referred to as *media-independent FEC*. If at least k out of n packets are correctly received, the underlying video information can be correctly decoded. Otherwise, none of the lost packets can be recovered by the

[4]Some MPEG-2 encoder chips automatically produce concealment motion vectors for all intra-coded macroblocks.

receiver. Hence the packet loss pattern experienced at the video level is quite different from the loss pattern observed on the global packet stream of the lossy channel.

The purpose of this section is to compute the video loss process parameters after FEC recovery. These parameters are the packet loss ratio π_v and the average burst length α_v. They directly drive the final video quality. The packet loss ratio π_v simply represents the ratio between lost and sent video packets. The average burst length α_v is the average length of consecutively lost video packets.

Several studies have been performed to compute the FEC efficiency or the probability for data to be recovered in case of loss [18–20]. The probability of recovery is simply given by the probability to have less than $n-k+1$ losses in a block of n packets. However this parameter do not bring enough information about the loss process after FEC recovery. To correctly model the video quality, at least two parameters (i.e, π_v and α_v) should be computed.

For sake of simplicity let first packets take binary values 0 and 1 for correctly received and lost packets respectively. Let us assume moreover that the loss process could be matched with a renewal error process. In other words, the lengths of consecutive inter-error intervals (also called gaps) are assumed to be independently and identically distributed.

Following the development of [18], let $p(i)$ denote the probability that a gap length is $i-1$, i.e., $p(i) = \Pr(0^{i-1}1|1)$, where 0^{i-1} is a shorthand for $i-1$ successive 0's. Similarly, let $P(i)$ denote the probability that at least $i-1$ 0's follow a given error, i.e., $P(i) = \Pr(0^{i-1}|1)$. It has to be noted that $P(i)$ could be written as

$$P(i) = \sum_{m=i}^{\infty} p(m). \tag{1}$$

The probability of the sequence $1\ 0^{i-1}$ is thus given by

$$\Pr(1\ 0^{i-1}) = \pi\ P(i). \tag{2}$$

Because of the independence among gap lengths of a renewal process order is irrelevant and the events $1\ 0^{i-1}$ and $0^{i-1}1$ are equiprobable. From these properties, the probability $R(m,n)$ that $m-1$ errors occur in the next $n-1$ bits following an error, could be easily computed by recurrence [18]. Thus,

$$R(m,n) = \begin{cases} P(n), & \text{for } m = 1 \text{ and } n \geq 1 \\ \sum_{i=1}^{n-m+1} p(i)\ R(m-1, n-i), & \text{for } 2 \leq m \leq n. \end{cases} \tag{3}$$

As mentioned earlier, this probability is not enough to model the video loss process. We will therefore introduce other probabilities to reach this goal. To this end, let $r(m,n)$

denote the probability that $m-1$ errors occur in the $n-1$ bits between two errors. It could also be computed by recurrence as

$$r(m,n) = \begin{cases} p(n), & \text{for } m=1 \text{ and } n \geq 1 \\ \sum_{i=1}^{n-m+1} p(i)\ r(m-1,n-i), & \text{for } 2 \leq m \leq n. \end{cases} \tag{4}$$

Finally, let $\bar{r}(m,n)$ represent the probability that $m-1$ errors occur in the $n-1$ following an error and preceding a 0:

$$\bar{r}(m,n) = R(m,n) - r(m,n). \tag{5}$$

From the above variables, the probability $q(i)$ to have a burst of length $i-1$ and the probability $Q(i)$ that at least $i-1$ 1's follow a zero could be easily computed. The dual of $R(m,n)$, $S(m,n)$ represents the probability to have $m-1$ 0's in the next $n-1$ bits following a 0. This probability could also be computed by recurrence by

$$S(m,n) = \begin{cases} Q(n), & \text{for } m=1 \text{ and } n \geq 1 \\ \sum_{i=1}^{n-m+1} q(i)\ S(m-1,n-i), & \text{for } 2 \leq m \leq n. \end{cases} \tag{6}$$

The video packet loss rate after FEC recovery is now straightforward. Two cases are considered with respect to the state of the last video packet of a FEC block. Its loss or its presence directly drives the loss process within the FEC block. By the renewal process properties, π_v could thus be computed as

$$\begin{aligned} \pi_v &= \frac{\pi}{k} \sum_{i=1}^{k} i\ R(i,k) \sum_{j=\lfloor n-k+1-i \rfloor}^{n-k} R(j+1,n-k+1) \\ &+ \frac{1-\pi}{k} \sum_{i=1}^{k-1} (k-i)\ S(i,k) \sum_{j=0}^{k-1-i} S(j+1,n-k+1), \end{aligned} \tag{7}$$

where the notation $\lfloor x \rfloor$ represents the positive part of x.

The average video burst length after FEC recovery, α_v could also be computed from the previous development. Since bursts of errors have not the same probability to start on any packet of the FEC block, each position has to be considered separately. The probability P_j that a burst starts at the j^{th} position of the FEC block is given by

$$P_j = \Pr(x_{j-1} = 0, x_j = 1), \tag{8}$$

where x_i denotes the binary state of the packet i (i.e., either lost or not after FEC recovery). The average length L_j of a burst starting at position j in a FEC block could then be written as

$$L_j = \frac{\sum_{l=1}^{\infty} l\ \Pr(x_{j-1}=0, x_j=1, \ldots, x_{j+l-1}=1, x_{j+l}=0)}{P_j}, \tag{9}$$

where only video packets are considered. Indeed FEC packets have to be skipped at the end of each FEC block. They obviously do not participate to the video burst length. The video average burst length α_v is finally given by the following probability-weighted summation

$$\alpha_v = \frac{\sum_{j=1}^{k} P_j \ L_j}{\sum_{j=1}^{k} P_j}. \tag{10}$$

Assume now that the channel loss process can be characterized by the Gilbert model [21–24]. Notice that our goal here is not to validate the Gilbert model, but rather to show how, given a channel model, the video loss pattern could be computed. The Gilbert model is a two-state Markovian model [25] with geometrically distributed residence times (see Figure 4). States 0 and 1 correspond respectively to the correct reception and loss of a packet. The transition rates p and q between the states control the lengths of the error bursts.

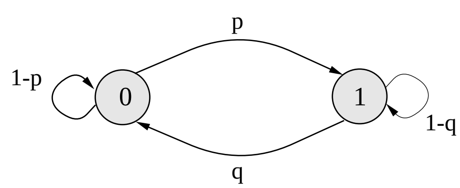
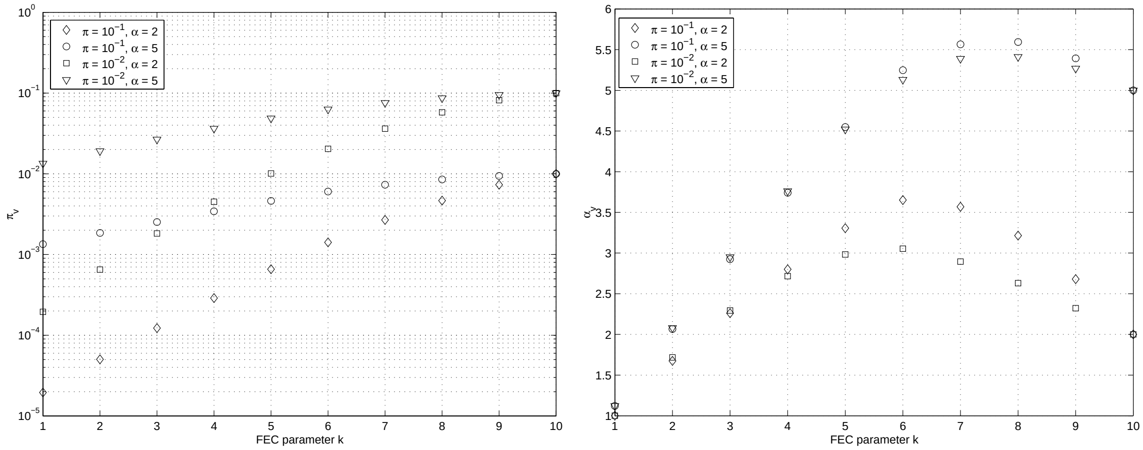


Fig. 4. Two-state Markov chain: Gilbert model.

The global packet loss ratio π corresponds to the stationary probability to be in the loss state: $\pi = \frac{p}{p+q}$. The average error burst length α is given by the average residence time in the loss state: $\alpha = \frac{1}{q}$. These loss process parameters π and α can be sensed through control protocols (e.g., RTCP) or delay measurements [26].

The loss patterns π_v and α_v could be easily computed in the case of a Gilbert-model loss process. They obviously depend on both the model parameters p and q and the FEC parameters k and n. Details of the computation are given in the Appendix A. For the remaining of the development, the parameters (π_v, α_v) will be written as $\pi_v(k, n)$ and $\alpha_v(k, n)$ to explicit their dependence on k and n.

Figures 5 (a) and (b) represents the evolution of the video loss parameters π_v and α_v for different network loss patterns. It is shown that π_v increases with k as the amount of protection decreases for a given n. Also FEC protection becomes less efficient for bursty loss traffic (large α values) for a given π. Moreover the average length of lost video packets clearly exhibits a maximum around $k = \frac{n}{2}$. This maximum can be explained as follows. When the amount of protection is very large, α_v stays close to k. When the amount of protection decreases, the video loss pattern gets closer to the channel loss pattern. In between there is an maximum which is less pronounced for bursty process. These behaviors still hold for different FEC block lengths.



(a) (b)

Fig. 5. Evolution of π_v and α_v versus the number of video packets k in a FEC block of length $n = 10$.

IV. MPEG-2 perceptual distortion-rate function

Several studies have already been conducted on the analysis of the rate-distortion curve for MPEG codecs [27–29]. They lead to the conclusion that the distortion evolves somehow exponentially with the decreasing source rate [30]. However, video distortion was measured by means of pure mathematical metrics such as the MSE.

Figure 6 shows that the perceptual video distortion varies exponentially with the quantizer scale factor (i.e., $MQUANT$). The TV-resolution video sequence was encoded in open-loop VBR mode (OL-VBR) using a TM-5 MPEG-2 video coder [31]. The video distortion was measured by means of the PDM tool [32], which proved to behave consistently with human judgments. This tool relies on a model of the human visual system (HVS) [33].

Moreover, the average source rate R_S is also evolving exponentially with the $MQUANT$ [34].

Therefore, the source perceptual distortion-rate function can be expressed as (see Figure 7):

$$D_S(R_S) = \chi_S \; R_S^{\xi_S}, \tag{11}$$

where the parameters χ_R and ξ_R are related to the encoding complexity of the set of frames under consideration. Details of the interpretation of Eq. (11) can be found in [35]. This relation holds at the sequence, group of pictures and even frame level.

Figure 7 exhibits an important, though trivial, behavior: for MPEG-2 source rates below 10 Mbps, a small increase in source rate leads to a great decrease in perceptual distortion.

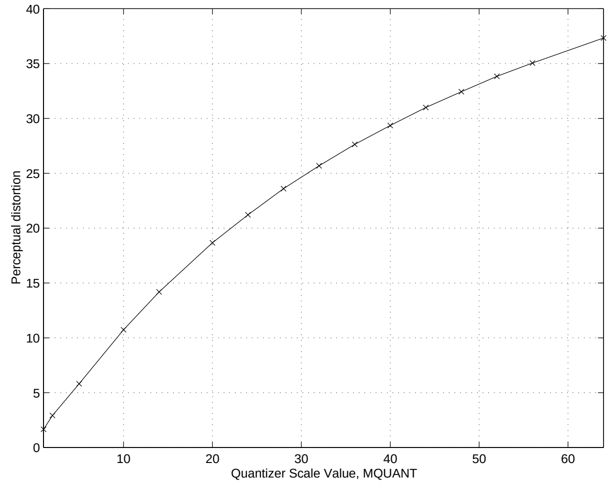


Fig. 6. Perceptual distortion as a function of the quantizer scale MQUANT.

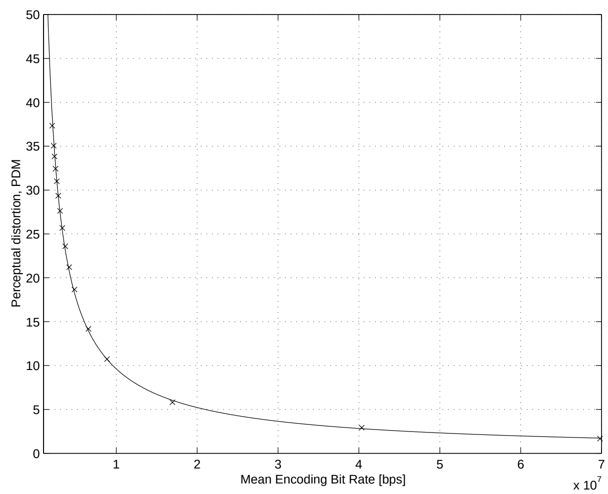


Fig. 7. Perceptual distortion as a function of the mean encoding bit rate. Fitting parameters of Eq. (11): $\chi_R = 14.59\ 10^6$ and $\xi_R = -0.883$.

V. Perceptual Distortion under Different Loss Patterns

We now investigate the impact of different packet loss patterns onto video distortion.

In MPEG-2 streaming over IP networks, a macroblock may be damaged in any of the three following cases:

- (i) it belongs to an RTP packet that has been lost during transmission,
- (ii) it belongs to a slice that has been affected by a packet loss (*spatial propagation*),
- (iii) it is temporally dependent on a damaged area of a previous reference frame (*temporal propagation*).

Intuitively speaking the perceptual distortion of the received video is in average proportional to the number of lost macroblocks, hence to the number of lost pixels. To emphasize this, an MPEG-2 transmission system has been simulated using a 400 frames long sequence and a very large set of network loss patterns. Figure 8 shows the video distortion as a function of the number of lost pixels. The linear proportionality of this result has a correlation factor lying around 0.992. Video distortion is thus directly related to the number of spatially lost macroblocks. Equivalently, the distortion is directly proportional to the number of spatially lost pixels, since macroblocks can only be entirely lost.

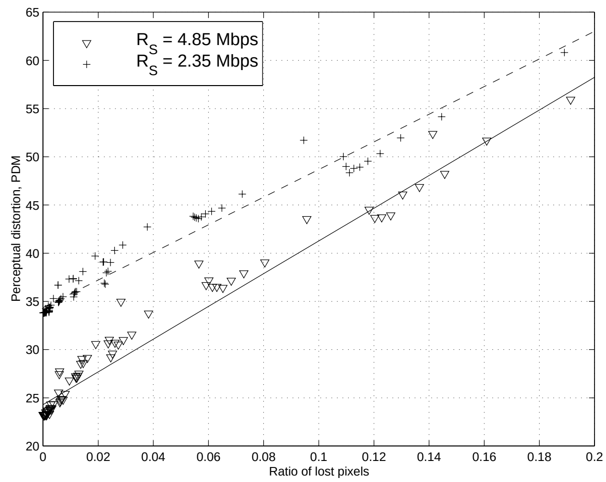


Fig. 8. Perceptual distortion versus percentage of spatially lost pixels for several loss patterns (source rate $R_S \in \{4.85, 2.35\}$ Mbps).

Hence the perceptual distortion is driven by spatial error propagation. Temporal error propagation is clearly a direct consequence of the lost macroblocks, hence pixels.

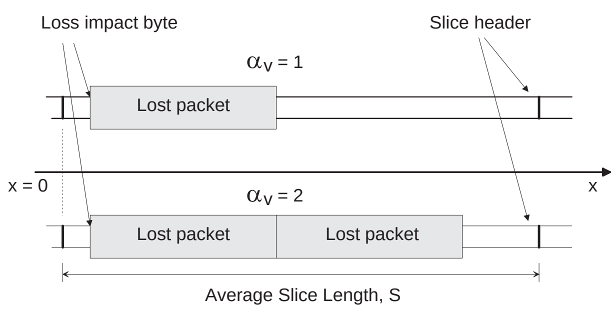


Fig. 9. Spatial error propagation under $\alpha_v = \{1, 2\}$.

Furthermore, notice that the loss of a single packet could be enough to lose the entire corresponding slice. As an example, Figure 9 shows how data loss propagates spatially for two different loss patterns (i.e., $\alpha_v = 1$ and $\alpha_v = 2$). Since an entire slice can be fully lost with the loss of a single packet, it is desirable that a burst of lost packets damage the same slice. Accordingly, under a given packet loss ratio, bursts of lost packets damage fewer pixels than individual lost packets. Thus the average number of lost bytes (equivalently, the average number of spatially damaged pixels) decreases exponentially with α_v (see Appendix B). Thus we draw the important conclusion that a uniform loss process corresponds to the worst case with respect to the amount of degradation due to data loss. The probability P_l for pixels to be lost is given by (Eq. (37), Appendix B)

$$P_l = \frac{\overline{L}}{R_S} = \pi_v + \frac{R_S \; \pi_v}{2 \; P \; N_S \alpha_v}, \tag{12}$$

where P represents the packet size (i.e., generally 184 bytes of MPEG-2 video data) and N_S represents the average number of MPEG-2 slices per second.

Now the effect of network losses has to be analyzed in terms of distortion. It has been shown that the perceptual distortion is proportional to the number of spatially lost pixels (see Fig. 8). Showing the dependence of π_v and α_v from k and n Eq. (12) leads to the distortion D_L due to data loss:

$$D_L(k,n) = \chi_L \; \left(\pi_v(k,n) + \frac{R_S \; \pi_v(k,n)}{2 \; P \; N_S \alpha_v(k,n)} \right), \tag{13}$$

where χ_L is a constant depending on the spatio-temporal complexity of the sequence and the error concealment scheme. The residual loss process parameters π_v and α_v directly depends on the FEC algorithm and the global loss process (See Section III).

The variation of $D_L(k,n)$ on π_v and α_v as given by Eq. (13) is shown in Figure 10, for a given source rate. It can be seen that the distortion indeed increases linearly with π_v. Also, under a given π_v, the distortion decreases exponentially with α_v. This behavior has also been observed with the MSE distortion metric.

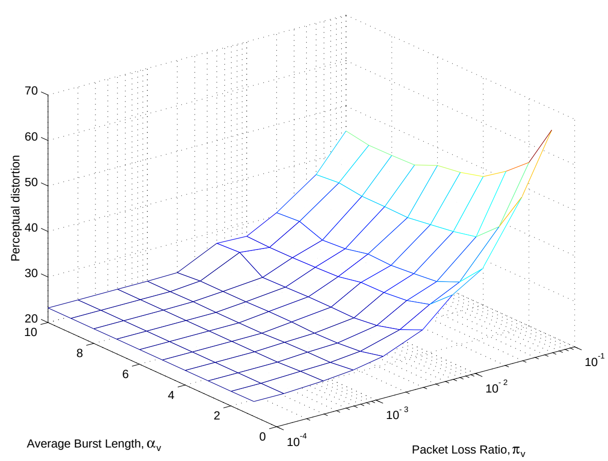


Fig. 10. Perceptual distortion versus the video loss process parameters (source rate $R_S = 4.75$ Mbit/s)

Finally, the mapping between the distortion $D_L(k,n)$ and the number of spatially lost macroblocks $\overline{L_p}$ (i.e., the value of χ_L) is almost independent of the source rate [35]. In the range of MPEG-2 target rates (i.e., 3.5 to 15 Mbps), the relative distortion of concealed areas is similar in low and high source rates. The quality of a reconstructed areas through error concealment depends on the accuracy of surrounding video elements, and thus on the source rate. For larger source rate, the reconstructed areas are better approximation of the original data. In the same time the quality of the lossless transmission also increases with the source rate. Hence the relative channel distortion is assumed to be independent of the source rate.

VI. Joint Source/Channel Perceptual Distortion

We now address the initial problem stated in Section I namely to find the optimal $R_S(t)$ and $R_{FEC}(t)$ at time t that *minimize* the end-to-end video distortion *under* the constraint $R_S(t) + R_{FEC}(t) \leq R(t)$ *given* a set of video and network parameters. The total bit rate $R(t)$ is one of the inputs to our system. It must be noted that this input may very well be adjusted to conform to any network policy (e.g., using the flow control algorithm of TCP [36]). Our rate allocation scheme then optimally uses the entire available rate $R(t)$.

By the FEC scheme structure, the source rate is expressed as $R_S(t) = \frac{k}{n}\ R(t)$ and the rate of FEC packets as $R_{FEC}(t) = \frac{n-k}{n} R(t)$ (see Figure 3). Recall that the distortion D_L as given by Eq. (13) represents the average distortion between a losslessly and packet-lossily transmitted versions of the same MPEG-2 bit stream. We need however a total end-to-end distortion measure D between the original and received video. Consider now only the video elements (e.g., macroblocks) that are lost but replaced by error concealment at the receiver. Let $\widetilde{D_L}$ be the average distortion between these elements and their original version. The end-to-end average distortion can then be written as

$$D = D_S\ (1 - P_l) + \widetilde{D_L}\ P_l = D_S + P_l\ (\widetilde{D_L} - D_S), \tag{14}$$

where P_l represents the average probability for a video element (e.g macroblock) to be lost. Eq. (14) holds for MSE-like distortion metrics. Assume that it is also verified in average for perceptual distortion. In this case, we can interpret the second-term in the righthand side of Eq. (14) as the average distortion due to data loss

$$D_L = P_l\ (\tilde{D}_L - D_S), \tag{15}$$

where D_L is directly related to the number of lost pixels from Eq. (13). Notice that Eq. (15) holds thanks to our definition of the distortion D_L.

Our problem becomes now the following: At time t, find the optimal FEC scheme parameters (k, n) that minmize the end-to-end distortion:

$$\min_{k \leq n \leq N} D(n,k) = \chi_S\ \left(\frac{k}{n}\ R(t)\right)^{\xi_S} + \chi_L\ \pi_v(k,n)\ \left(1 + \frac{k\ R(t)}{2\ n\ P\ N_S\ \alpha_v(k,n)}\right) \tag{16}$$

given the total bandwidth $R(t)$, the channel state (i.e., p and q in the Gilbert model) and the scene-dependent parameters (χ_S, ξ_S, χ_L). The constraint $n \leq N$ simply imposes a maximum reconstruction time for the FEC decoding.

Since the parameters (k, n) only take integer values, Eq. (16) can be solved easily through numerical methods. The optimal values of k and n then lead to the optimal rate distribution between source and FEC rate. In the same time, they define the optimal media-independent FEC algorithm in terms of the end-to-end distortion.

VII. Experimental Results

The optimal FEC parameters are given in Table I for several network conditions and video scenes of different spatio-temporal complexities. The FEC reconstruction delay is set to approximately 6.5 ms (i.e, $N = 20$ at 4 Mbps and $N = 30$ at 6 Mbps). Optimal FEC parameters and hence rate distribution are then numerically computed from Eq. 16. It is

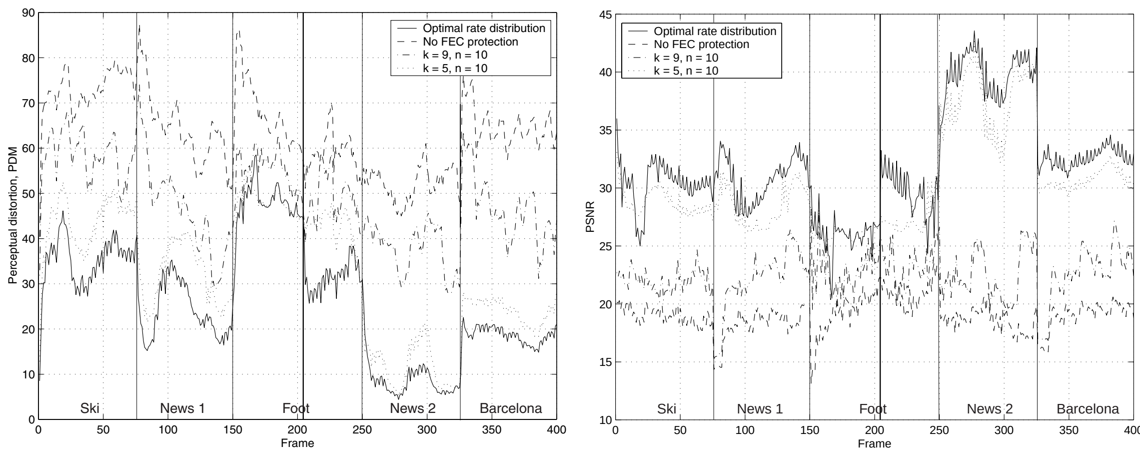
		$R = 4\ Mbps$				$R = 6\ Mbps$			
		Ski		Foot		Foot		News	
PLR	ABL	k	n	k	n	k	n	k	n
0.1	1	14	18	16	20	23	29	22	29
	2	14	20	15	20	22	30	19	30
0.01	1	19	20	19	20	28	30	28	30
	2	1	1	1	1	29	30	25	30
0.001	1	1	1	1	1	1	1	29	30
	2	1	1	1	1	1	1	1	1

TABLE I

Optimal FEC parameters for several transmissions conditions and scenes.

shown that the FEC rate (i.e., $\frac{n-k}{n}$) decreases when the global loss ratio (PLR) decreases. Moreover the required FEC rate is lower for the Foot sequence (i.e., high spatio-temporal complexity) than for the News video scene (i.e., low spatio-temporal complexity), at least for $PLR \leq 0.1$. This intuitive result clearly exhibits the joint role of both the source and FEC rates onto the end-to-end distortion. The evolution of the optimal parameters with the increasing global loss burstiness (ABL) is less straightforward. They indeed result from a tradeoff between the decreasing FEC efficiency and decreasing error propagation effect.

The temporal evolution of the distortion through a five-scene sequence is given in Figures 11(a) and 12(a). The distortion averaged over sliding windows through Minkowski summation is compared to the one obtained from classical FEC schemes. Thanks to its



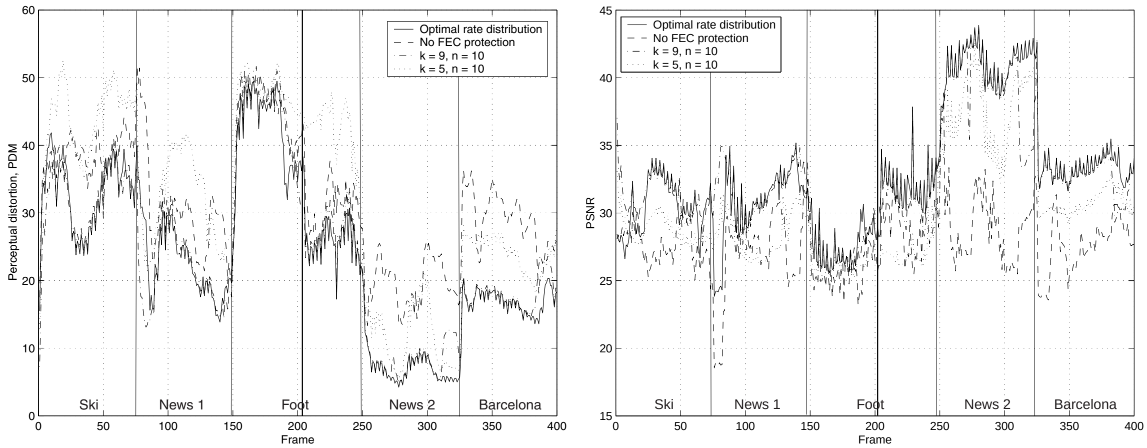
(a) (b)

Fig. 11. Perceptual distortion (a) and PSNR (b) vs frame number for optimal rate distribution scheme and classical FEC schemes. The sequence is composed of 5 different scenes. The available rate R is set to 4 Mbps for the 204 first frames, and to 6 Mbps for the 196 last frames. The global loss process parameters are set to $(PLR, ABL) = (.1, 1)$.

adaptivity features, our rate distribution algorithm outperforms the common schemes, as reported also by the PSNR evolution (see Figures 11(b) and 12(b)). The proposed algorithm indeed adapts to the scene complexity, to the available bandwidth and to the network conditions. It has to be noted that large loss ratio values have been chosen since the length of the sequence is relatively short. However, even if these values seem relatively high compared to usual mean ratios on today's network, they are likely to happen during small time intervals.

VIII. Conclusions

In this paper we considered a joint source and channel coding problem. More specifically, we proposed a solution to the problem of optimal rate distribution between MPEG-2 and media-independent FEC. The optimality has been defined in terms of minimal end-to-end perceptual video distortion. The efficiency of a media-independent FEC algorithm has been studied first. The residual video loss patterns after FEC recovery have been computed for a Gilbert-model global loss process. The exponential source perceptual distortion-rate function has then be derived from empirical results. Finally, the distortion due to loss has been shown to be directly proportional to the number of lost pixels. From this set of equations, the optimal rate distribution, as well as the optimal FEC scheme



(a) (b)

Fig. 12. Perceptual distortion (a) and PSNR (b) vs frame number for optimal rate distribution scheme and classical FEC schemes. The sequence is composed of 5 different scenes. The available rate R is set to 4 Mbps for the 204 first frames, and to 6 Mbps for the 196 last frames. The global loss process parameters are set to $(PLR, ABL) = (.01, 1)$.

are obtained by solving a simple optimization problem. Finally, the proposed allocation scheme has been shown to outperform classical FEC schemes due to its adaptivity to the video scene, to the available bandwidth and to the network state.

The same study could be applied to other encoding schemes (e.g., H.263, MPEG-4), although in this case the loss spatial propagation is reduced.

IX. Acknowledgments

The authors are particularly grateful to Professor Murat Kunt from the Signal Processing Laboratory for his relevant comments and remarks on the paper.

Appendix

I. FEC recovery in a Gilbert-model loss process

For a Gilbert loss process the following relations hold

$$p(i) = \begin{cases} 1 - q, & \text{if } i = 1 \\ q\ (1-p)^{i-2}\ p, & \text{otherwise.} \end{cases} \tag{17}$$

$$P(i) = \begin{cases} 1, & \text{if } i = 1 \\ q\ (1-p)^{i-2}, & \text{otherwise.} \end{cases} \tag{18}$$

$$q(i) = \begin{cases} 1-p, & \text{if } i = 1 \\ p\ (1-q)^{i-2}\ q, & \text{otherwise.} \end{cases} \tag{19}$$

$$Q(i) = \begin{cases} 1, & \text{if } i = 1 \\ p\ (1-q)^{i-2}, & \text{otherwise.} \end{cases} \tag{20}$$

The probabilities $R(m,n)$, $S(m,n)$, $r(m,n)$ and $\bar{r}(m,n)$ could be computed by recurrence from Eqs. (3), (6), (4) and (5) respectively.

Let now R_n the probability that no packet is lost in a FEC block given that the first packet is erased. It could be written as

$$R_n = \sum_{i=0}^{n-k-1} R(i+1,n). \tag{21}$$

The probability for a burst to start on j^{th} packet of a FEC block (i.e., $1 \leq j \leq n$) is thus given by

$$P_j = \begin{cases} \pi\ \ (q + (1-q)\ R_n)\ \sum_{i=n-k}^{n-1} R(i+1,n), & \text{if } j = 1 \\ \pi\ \sum_{i=0}^{n-j} \sum_{x=0}^{k-n+j+i-2} q\ S(x+1,j-1)\ R(i+1,n-j+1), & \text{if } j > 1. \end{cases} \tag{22}$$

The average length of bursts of video packets, α_v, has now to exclude redundancy packets. Let first f denote the number of FEC blocks transitions along the burst. It could be written as

$$f = \text{floor}\left(\frac{l+j-2}{k}\right). \tag{23}$$

Let now z denote the position in a FEC block of the first video packet following a burst of length l:

$$z = \begin{cases} k, & \text{if } (l+j-1) \bmod k = 0 \\ (l+j-1) \bmod k, & \text{otherwise.} \end{cases} \tag{24}$$

Moreover let r_{n-k} and $\bar{r}_{n-k}$ denote the probability to lose all the video packets of a FEC block and that the first video packet of the next FEC block is either erased or not (before FEC recovery). These probabilities could be written as

$$r_{n-k} = \sum_{\lfloor n-2k+1 \rfloor}^{n-k} r(i+1,n-k+1), \tag{25}$$

and

$$\bar{r}_{n-k} = \sum_{\lfloor n-2k+1 \rfloor}^{n-k} \bar{r}(i+1,n-k+1), \tag{26}$$

where the notation $\lfloor x \rfloor$ represents the positive part of x. Similarly r^j_{n-k} and $\bar{r}^j_{n-k}$ denote the probability that the last video packet (i.e., the packet k) of the first FEC block is lost, and the first video packet of the second block is respectively erased or not (before FEC recovery). It is assumed that all video packet between packet j and k are lost. These probabilities could hence be written as

$$r^j_{n-k} = \sum_{i=0}^{n-k} r(i+1, n-k+1) \sum_{x=0}^{2k-n+i-2} S(x+1, j-1), \tag{27}$$

and

$$\bar{r}^j_{n-k} = \sum_{i=0}^{n-k} \bar{r}(i+1, n-k+1) \sum_{x=0}^{2k-n+i-2} S(x+1, j-1). \tag{28}$$

The average length of a burst starting on the j^{th} video packet of a FEC block is given by

$$L_j = \sum_{l=0}^{\infty} l\ (1-q)^{l-f-1}\ c_j(l), \tag{29}$$

where $c_j(l)$ is the probability to have a video packet burst of length l starting with j. The conditional probability $c_1(l)$ for bursts starting with the first video packet of the FEC block (i.e., $j = 1$) could be written as

$$c_1(l) = \begin{cases} \frac{\pi\ (q+(1-q)\ R_n)}{P_1}\ (r_{n-k}\ R_n + \bar{r}_{n-k})\ {r_{n-k}}^f, & \text{if } z = n \\ \frac{\pi\ (q+(1-q)\ R_n)}{P_1}\ q\ \sum_{i=0}^{k-2} S(i+1, n-z)\ {r_{n-k}}^f, & \text{otherwise.} \end{cases} \tag{30}$$

For $2 \leq j \leq k$, c_j could be written as

$$c_j(l) = \begin{cases} \frac{\pi\ q}{P_j}\ \sum_{i=0}^{n-z-1} \sum_{x=0}^{k-3-i} S(x+1, j-1)\ q\ S(i+1, n-z), & \text{if } l+j-1 < k \\ \frac{\pi\ q}{P_j}\ (r^j_{n-k}\ R_n + \bar{r}^j_{n-k}), & \text{if } z = k \text{ and } f = 0 \\ \frac{\pi\ q}{P_j}\ (r_{n-k}\ R_n + \bar{r}_{n-k})\ r^j_{n-k}\ {r_{n-k}}^f, & \text{if } z = k \text{ and } f \neq 0 \\ \frac{\pi\ q}{P_j}\ q \sum_{i=0}^{k-2} S(i+1, n-z)\ r^j_{n-k}\ {r_{n-k}}^{\lfloor f-1 \rfloor}, & \text{otherwise.} \end{cases} \tag{31}$$

Finally, α_v is given by

$$\alpha_v = \frac{\sum_{j=1}^{k} P_j\ L_j}{\sum_{j=1}^{k} P_j} \tag{32}$$

II. Number of spatially lost pixels

Let x denote the byte position where the loss starts. Let then σ represent the integer number of slices damaged by the loss. If P and S respectively represent the size of the transmission packet and the average slice length, σ has to verify

$$(\sigma - 1)\ S < \alpha_v\ P \leq \sigma\ S. \tag{33}$$

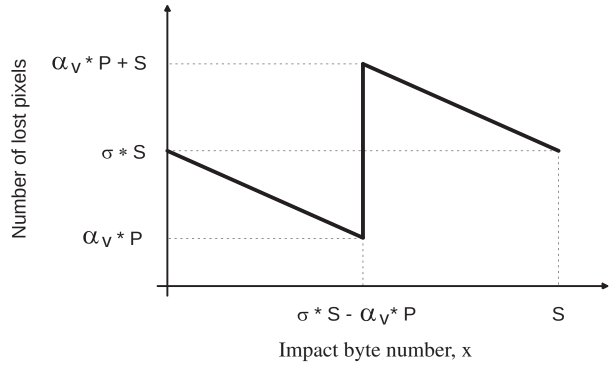


Fig. 13. Number of damaged pixels versus the impact byte in the stream.

From Figure 9, it can be deduced that the number of lost pixels follows a piecewise linear function versus the impact byte, as represented in the Figure 13. The average number of lost video bytes per burst is therefore given by

$$\overline{B} = \alpha_v \ P + \frac{S}{2} = \alpha_v \ P + \frac{R_S}{2 \ N_S}, \tag{34}$$

where R_S represents the source rate. N_S denote the number of slices per time unit. The mean rate of packet loss bursts is then given by

$$\overline{R_B} = \frac{R_S \ \pi_v}{P \ \alpha_v}. \tag{35}$$

Finally, from Eq. (34) and (35), the rate of lost video bytes, $\overline{L}$, is simply given by

$$\overline{L} = \overline{B} \ \overline{R_B} = R_S \ \pi_v + \frac{{R_S}^2 \ \pi_v}{2 \ P \ N_S \alpha_v}. \tag{36}$$

Since in average the number of lost pixels is proportional to the number of lost video bytes, the ratio of lost pixels is equivalent to

$$P_l = \frac{\overline{L}}{R_S} = \pi_v + \frac{R_S \ \pi_v}{2 \ P \ N_S \alpha_v}. \tag{37}$$

It has to be noticed that P_l represents also the probability for a pixel to be lost.

To show the dependence of P_l on α_v, Figure 14 represents the number of spatially damaged pixels versus the video average burst length and the packet loss ratios π_v. The number of lost pixels is clearly decreasing in $1/\alpha_v$. Moreover, the phenomenon is more visible for low α_v values (i.e., typical values). Finally, it has to be noted that the influence of the Average Burst Length decreases when the encoding bit rate decreases, since the slice size decreases.

Finally, it has to be noticed that the computed value of P_l is very close to experimental data. The correlation factor computed between simulation results and analytic value from Eq. (37) lies around .995 for a large set of source rate.

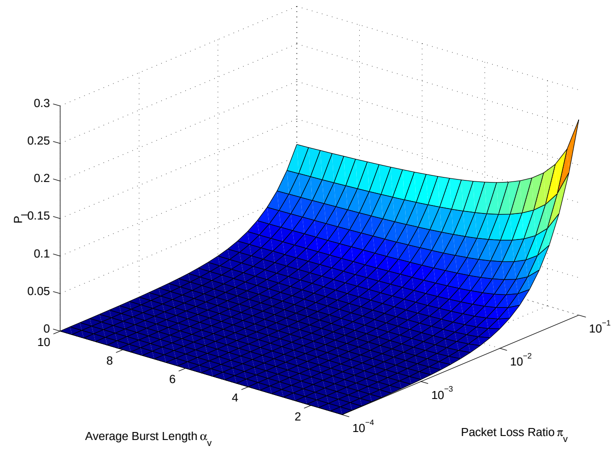


Fig. 14. Number of lost pixels versus the Average Burst Length α_v and the Packet Loss Ratio π_v (source rate $R_S = 4.75$ Mbit/s).

References

[1] Verscheure O., Garcia Adanez X., Karlsson G. and Hubaux J.-P., "User-Oriented QoS in Packet Video Delivery," *IEEE Network Magazine*, vol. 12, no. 6, pp. 12–21, November/December 1998.

[2] Richardson I. E. G. and Riley M. J., "Improving the error tolerance of MPEG video by varying slice size," *Signal Proceesing*, vol. 46, pp. 369–72, 1995.

[3] Frossard P. and Verscheure O., "MPEG-2 Video over Lossy Packet Networks: Perceptual Syntactic Information Coding," in *SPIE's International Symposium on Voice, Video, and Data Communications*, Boston, November 1998, vol. 3528, pp. 113–123.

[4] Zhang J., Frater M.R., Arnold J.F. and Percival T.M., "MPEG 2 Video Services for Wireless ATM Networks," *Journal on Selected Areas in Communications*, vol. 15, no. 1, pp. 119–128, January 1997.

[5] Horn U., Stuhlmuller K., Link M. and Girod B., "Robust Internet Video Transmission Based on Scalable Coding and Unequal Error Protection," *Signal Processing: Image Communications*, vol. 15, no. 1-2, pp. 77–94, 1999.

[6] McCanne S., Vetterli M. and Jacobson V., "Low-complexity Video Coding for Receiver-driven Layered Multicast," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 983–1001, August 1997.

[7] Perkins C., Hodson O. and Hardman V., "A Survey of Packet-Loss Recovery Techniques for Streaming Audio," *IEEE Network Magazine*, vol. 12, no. 5, pp. 40–48, September 1998.

[8] McAuley A.J., "Reliable Broadband Communications Using a Burst Erasure Correcting Code," in *ACM SIGCOMM'90*, Philadelphia, USA, September 1990, pp. 297–306.

[9] IETF, Audio/Video Transport Working Group, "RTP: A Transport Protocol for Real-Time Applications," Standard Track RFC:1889, IETF, January 1996.

[10] IETF, Audio/Video Transport Working Group, "RTP Payload Format for MPEG1/MPEG2 Video," Internet Draft draft-ietf-avt-mpeg-new-01.txt, IETF, June 1997.

[11] Wang Y. and Zhu Q.-F., "Error Control and Concealment for Video Communication: A Review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[12] Lopez Fernandez C., Basso A. and Hubaux J.-P., "Error concealment and early resynchronization techniques for MPEG-2 video streams damaged by transmission over ATM networks," in *Proceedings of the SPIE*, 1996, vol. 2668, pp. 372–383.

[13] Rabin M.O., "Efficient Dispersal of Information for Security, Load Balancing and Fault Tolerance," *Journal of the ACM*, vol. 36, no. 2, pp. 335–348, April 1989.

[14] Rizzo L., "Effective Erasure Codes for Reliable Computer Communication Protocols," *ACM Computer Communication Review*, vol. 27, no. 2, pp. 24–36, April 1997.

[15] Lin S. and Costello D.J.jr, *Error control coding: fundamentals and applications*, Prentice-Hall, 1983.

[16] Shacham N. and McKenney P., "Packet recovery in High-Speed Networks Using Coding and Buffer Management," in *Proceedings of the IEEE INFOCOM Conference*, Los Alamitos, CA, 1990, vol. 1, pp. 124–131.

[17] Blahut R.E., *Theory and Practice of Error Control Codes*, Addison Wesley, 1983.

[18] Elliott E.O., "A Model of the Switched Telephone Network for Data Communications," *The Bell System Technical Journal*, vol. 44, no. 1, pp. 89–109, January 1965.

[19] Adoul J.-P.A., "Error Intervals and Cluster Density in Channel Modeling," *IEEE Transactions on Information Theory*, vol. 20, pp. 125–129, January 1974.

[20] Kanal L.N. and Sastry A.R.K., "Models for Channels with Memory and Their Applications to Error Control," *Proceedings of the IEEE*, vol. 66, no. 7, pp. 724–744, July 1978.

[21] Bolot J., Fosse-Parisis S. and Towsley D., "Adaptive FEC-Based Error Control for Interactive Audio in the Internet," in *Proceedings of the IEEE INFOCOM*, 1999.

[22] Yajnik M., Kurose J. and Towsley D., "Packet Loss Correlation in the MBone Multicast Network: Experimental Measurements and Markov Chain Models," Tech. Rep. 95-115, Department of Computer Science, University of Massachusetts at Amherst, 1995.

[23] Paxson V., "End-to-End Internet Packet Dynamics," *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 4, pp. 139–152, October 1997.

[24] Jeruchim M.C., Balaban P. and Shanmugan K.S., *Simulation of communication systems*, Plenum Press, 1992.

[25] Gilbert E.N., "Capacity of a Burst-Noise Channel," *The Bell System Technical Journal*, pp. 1253–1265, September 1960.

[26] Paxson V., Almes G., Mahdavi J. and Mathis M., "Framework for IP Performance Metrics," Tech. Rep. RFC 2330, IETF, May 1998.

[27] Ramchandran K., Ortega A. and Vetterli M., "Bit Allocation for Dependent Quantization with Applications to Multiresolution and MPEG Video Coders," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 533–545, September 1994.

[28] Ding W. and Liu B., "Rate Control of MPEG Video Coding and Recording by Rate-Quantization Modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 1, pp. 12–20, February 1996.

[29] Chiang T. and Zhang Y.-Q., "A New Rate Control Scheme Using Quadratic Rate Distortion Model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246–250, February 1997.

[30] Lin L.-J. and Ortega A., "Bit-Rate Control Using Piecewise Approximated Rate-Distortion Characteristics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 446–459, August 1998.

[31] Fogg C., "mpeg2encode/mpeg2decode," in *MPEG Software Simulation Group*, 1996.

[32] Winkler S., "A Perceptual Distortion Metric for Digital Color Video," in *Proceedings of SPIE Human Vision and Electronic Imaging*, San Jose, CA, Jan. 23–29, 1999, vol. 3644.

[33] van den Branden Lambrecht C. J., *Perceptual Models and Architectures for Video Coding Applications*, Ph.D. thesis, Swiss Federal Institute of Technology, CH 1015 Lausanne, Switzerland, 1996.

[34] Sakazawa S., Takishima Y., Wada M. and Hatori Y., "Coding Control Scheme for a Multi-Encoder System," in *Proceedings of the 7th International Workshop on Packet Video*, March 1996, pp. 83–88.

[35] Verscheure O., Frossard P. and Hamdi M., "User-Oriented QoS Analysis in MPEG-2 Video Delivery," *Journal of Real-Time Imaging*, vol. 5, no. 5, pp. 305–314, October 1999.

[36] Jacobson V., "Congestion Avoidance and Control," in *Proceedings of the ACM SIGCOMM*, 1988.