JOINT SOURCE/FEC RATE SELECTION FOR OPTIMAL MPEG-2 VIDEO DELIVERY

Pascal Frossard

Signal Processing Laboratory
Swiss Federal Institute of Technology
Lausanne, Switzerland
pascal.frossard@epfl.ch

Olivier Verscheure

IBM T.J. Watson Research Center
New York, USA
ov1@us.ibm.com

ABSTRACT

This paper deals with the optimal allocation of MPEG-2 encoding and media-independent 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 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 schemes, thanks to its adaptivity to the scene complexity, to the available bandwidth and to the network conditions.

1. INTRODUCTION

Streamed digital video has already begun its penetration in the market. 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.

Interactive video delivery can significantly be improved by providing sender-side mechanisms [1], like forward error correction (FEC) mechanisms. 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. That is, k video packets are protected by $n-k$ FEC packets within a total of n, regardless of the underlying video sensitivity [2].

Clearly, under a given channel rate, the addition of FEC packets reduces the available rate for source coding. The optimal allocation between source coding rate and FEC rate under a given total bit rate is the subject 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. 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[1] *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 2 briefly describes the sensitivity of MPEG-2 video to data loss. Section 3 analyzes the FEC efficiency when the global loss process is modeled by a two-state markov chain. The source perceptual distortion-rate function is empirically computed in Section 4. Section 5 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 6. Section 7 shows how our algorithm behaves under different conditions, and compares the end-to-end distortion to classical FEC schemes. Finally, concluding remarks are given in Section 8.

2. MPEG-2 SENSITIVITY TO DATA LOSS

In MPEG-2, 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.

The error visibility may be dramatically reduced by means of error concealment techniques [3]. 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). The encoding process is therefore extended to include motion vectors for intra macroblocks also.

---

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

0-7803-6536-4/00/$10.00 (c) 2000 IEEE

3. 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 [4, 5] provide an efficient way to fight against losses, although the perfect recovery cannot be guaranteed.

Assume every block of k video packets is protected by $(n-k)$ FEC packets. 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 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 video loss process after FEC recovery can be represented by two major parameters. That is, 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.

These parameters are only dependent on the global loss process and the FEC parameters k and n. In the remainder of this paper, they will simply be noted $\pi_v(k,n)$ and $\alpha_v(k,n)$. The reader is referred to [6] for the equations along with the complete details of the computation of π_v and α_v when the loss process can be matched with a renewal error process.

4. MPEG-2 PERCEPTUAL DISTORTION-RATE FUNCTION

Several studies have already been conducted on the analysis of the rate-distortion curve for MPEG codecs [7, 8]. They lead to the conclusion that the distortion evolves somehow exponentially with the decreasing source rate [9]. However, video distortion was measured by means of pure mathematical metrics such as the MSE.

The perceptual video distortion as measured by means of the PDM tool [10] varies exponentially with the quantizer scale factor (i.e., $MQUANT$) [6]. The PDM tool relies on a model of the human visual system (HVS) [11]. Moreover, the average source rate R_S is also evolving exponentially with the $MQUANT$ [12, 13].

Therefore, the source perceptual distortion-rate function can be expressed as (see Figure 1):

$$D_S(R_S) = \chi_S \, R_S^{\xi_S}, \tag{1}$$

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. (1) can be found in [14]. This relation holds at the sequence, group of pictures and even frame level. Figure 1 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.

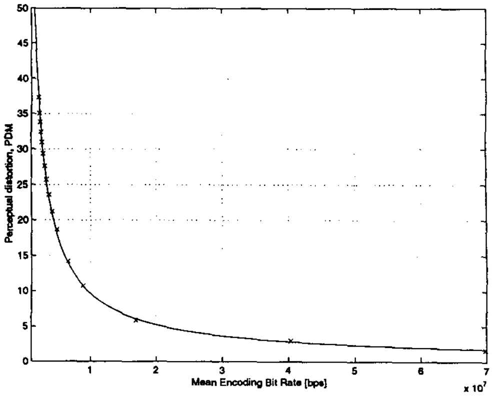

Figure 1: Perceptual distortion as a function of the mean encoding bit rate. Fitting parameters of Eq. (1): $\chi_R = 14.59\ 10^6$ and $\xi_R = -0.883$.

5. 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 sequence and a very large set of network loss patterns. The resulting linear proportionality 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.

Hence the perceptual distortion is driven by spatial error propagation. Temporal error propagation is clearly a direct consequence of the lost macroblocks, hence pixels. Furthermore, notice that the loss of a single packet could be enough to lose the entire corresponding slice. 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. 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 (please refer to [6] for the details):

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

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. Showing the dependence of π_v and α_v from k and n Eq. (2) leads to the distortion D_L due to data loss:

$$D_L(k,n) = \chi_L \left(\pi_v(k,n) + \frac{R_S\ \pi_v(k,n)}{2\ P\ N_S \alpha_v(k,n)} \right), \quad (3)$$

where χ_L is a constant depending on the spatio-temporal complexity of the sequence and the error concealment scheme.

The impact of π_v and α_v onto $D_L(k,n)$ for a given source rate is shown in Figure 2. The distortion D_L 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. 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 [14].

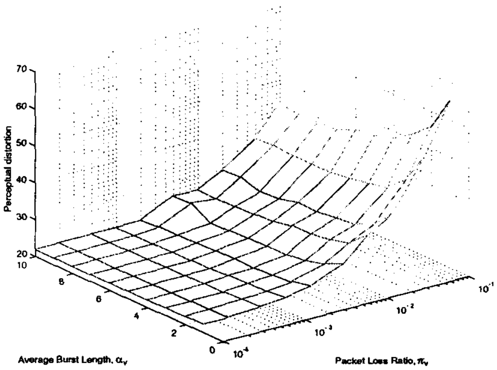

Figure 2: Perceptual distortion versus the video loss process parameters (source rate R_S = 4.75 Mbit/s)

6. JOINT SOURCE/CHANNEL PERCEPTUAL DISTORTION

We now address the initial problem. That is, 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.

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 $R_{FEC}(t) = \frac{n-k}{n} R(t)$. Recall that the distortion D_L as given by Eq. (3) represents the average distortion between the compressed video MPEG-2 sequence before and after transmission. However we need to come up with the end-to-end distortion D. 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

$$D = D_S\ (1 - P_l) + \widetilde{D_L}\ P_l = D_S + P_l\ (\widetilde{D_L} - D_S), \quad (4)$$

where P_l represents the average probability for a video element (e.g macroblock) to be lost. Eq. (4) 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. (4) as the average distortion due to data loss

$$D_L = P_l\ (\tilde{D_L} - D_S). \quad (5)$$

Notice that Eq. (5) holds thanks to the definition of the distortion D_L.

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

$$\begin{aligned} \min_{k \leq n \leq N} D(n,k) &= \chi_S \left(\frac{k}{n}\ R(t) \right)^{\xi_S} \quad (6) \\ &+ \chi_L\ \pi_v(k,n) \left(1 + \frac{k\ R(t)}{2\ n\ P\ N_S\ \alpha_v(k,n)} \right) \end{aligned}$$

given the total bandwidth $R(t)$, the channel state (i.e., p and q in the Gilbert model [6]) and the video 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. (6) 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 rates. In the same time, they define the optimal media-independent FEC algorithm in terms of the end-to-end distortion.

7. EXPERIMENTAL RESULTS

The optimal FEC parameters are given in Table 1 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. 6. It is 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

		$R = 4\ Mbps$				$R = 6\ Mbps$			
		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 1: Optimal FEC parameters for several transmissions conditions and scenes.

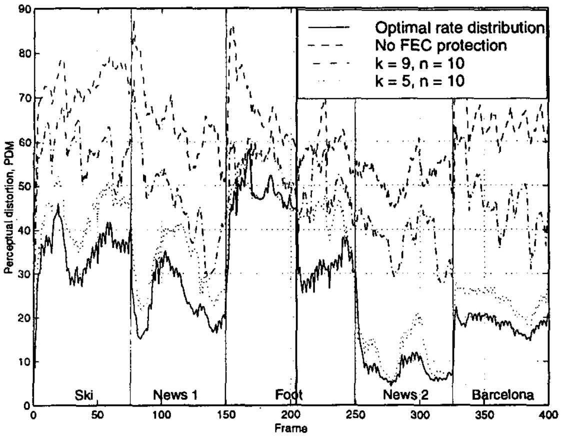

Figure 3: Perceptual distortion vs frame number for optimal rate distribution scheme and classical FEC schemes. The sequence is composed of 5 different scenes. The available rate R is set to 4 Mbps for the 204 first frames, and to 6 Mbps for the 196 last frames. The global loss process parameters are set to $(PLR, ABL) = (.1, 1)$.

the decreasing FEC efficiency and decreasing error propagation effect. The temporal evolution of the distortion through a five-scene sequence is given in Figure 3. The distortion averaged over sliding windows through Minkowski summation is compared to the one obtained from classical FEC schemes. Thanks to its adaptivity features, our rate distribution algorithm outperforms the common schemes. 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.

8. 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 media-independent FEC and MPEG-2. 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 discussed first. 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 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.

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

[3] 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.

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

[5] 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.

[6] Frossard P. and Verscheure O., "Joint Source/FEC Rate Selection for Quality-Optimal MPEG-2 Video Delivery," Tech. Rep. LTS 1999-04, Signal Processing Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland, November 1999, available on http://dewww.epfl.ch/ frossard/publications.html.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.