Published in Picture Coding Symposium, Berlin, September 1997

PERCEPTUAL MPEG-2 SYNTACTIC INFORMATION CODING: A PERFORMANCE STUDY BASED ON PSYCHOPHYSICS

Olivier Verscheure and Pascal Frossard

TCOM Laboratory, Telecommunication Services Group
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
E-Mail: {verscheure, frossard}@tcom.epfl.ch
URL : http://tcomwww.epfl.ch/{~verscheu, ~frossard}

ABSTRACT

This work addresses the optimization of TV-resolution MPEG-2 video streams to be transmitted over packet networks facing a given packet loss probability. Several studies have already been carried out on this specific problem [1, 2, 3]. The main contribution of this work resides in the design of a new flexible scene-complexity adaptive mechanism, namely the Perceptual Syntactic Information Coding (PSIC) mechanism. The PSIC adaptively modulates the number of macroblocks per slice in order to optimize the global video quality which depends on both the video encoding quality and the video degradations due to data loss during the transmission. We compare its performance against a fixed-length slice scheme to be transmitted over ATM networks using AAL-5. All the quality assessments are based on a perceptual quality metric, MPQM [1], which proved to behave consistently with human judgements [4]. Results show that the proposed mechanism behaves much better in terms of global perceptual video quality.

1. INTRODUCTION

The Asynchronous Transfer Mode (ATM) technology is reaching a certain level of maturity that permits its deployment in local as well as in wide area networks. One of the key points of such a technology is that it may provide statistical guarantees on performance. In other words, each user may agree with the network on a set of performance parameters (i.e., Quality of Service (QoS) parameters) which will have to be met.

Audiovisual applications (e.g., video conferencing, video on demand, teleteaching, etc.) are foreseen as one of the major users of such broadband networks. At the heart of this revolution is the digital compression of audio and video signals. The biggest advantage of compression resides in data rate reduction which results in reduction of transmission costs. The choice of the compression algorithm mostly depends on the available bandwidth or storage capacity and the features required by the application. The MPEG-2 [2] standard [5], a truly integrated audio-visual standard developed by the International Organization for Standards (ISO), is capable of compressing NTSC or PAL video into an average bit rate of 3 to 6 Mbits/s with a quality comparable to analog CATV [6]. However, a lot of work remains to be done to optimize these audiovisual applications so that they can be offered at attractive prices. In other words, the user expects an adequate perceptual audiovisual quality at the lowest possible cost. In the case of video transmission over packet networks, the service quality from the user viewpoint (i.e., the global perceptual video quality) results both from the video encoding quality and the degradations due to data loss during the transmission.

This work focuses on the optimization of TV-resolution MPEG-2 video streams to be transmitted over packet networks facing a given packet loss probability. The paper is organized as follows: Section 2 first introduces the MPEG-2 video and system standards and then briefly describes the impact of data loss onto the reconstructed video sequence. Section 3 presents the proposed syntactic information coding mechanism (PSIC). Sec. 3 first starts with some preliminary studies. Then, the PSIC mechanism is presented and, finally, its performance over ATM networks is analyzed. Concluding remarks are given in Sec. 4.

[1] MPQM stands for Moving Pictures Quality Metric

[2] MPEG stands for Moving Picture Experts Group

2. MPEG-2 TRANSMISSION OVER PACKET NETWORKS

2.1. MPEG-2 Overview

An MPEG-2 video stream [7] is highly hierarchically structured as illustrated in Fig. 1. The stream consists of a sequence composed of several pictures. The MPEG-2 video standard defines three different types of pictures : intra-coded (I-), predicted (P-) and bidirectional (B-) pictures. The use of these three picture types allows MPEG-2 to be robust (I-pictures provide error propagation reset points) and efficient (B- and P-pictures allow a good overall compression ratio). Each picture is composed of slices which are, by definition, a 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.

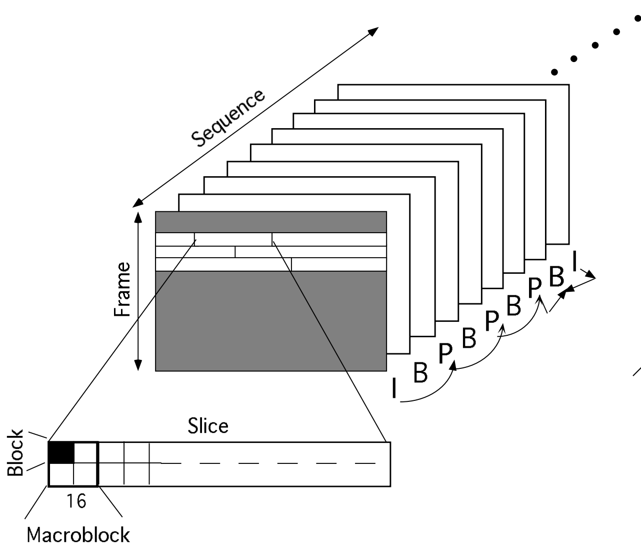

Figure 1: *MPEG-2 video structure.*

Before being transmitted, such a video stream goes through the MPEG-2 Transport Stream (TS) layer. Basically, the stream is first segmented into variable-length Packetized Elementary Stream packets and then subdivided into fixed-length TS packets. Moreover, it is worth noting that a non-encoded header (i.e., syntactic information) is inserted before each of the following information elements: sequence, Group of Pictures (GoP), picture, slice, TS and PES.

2.2. MPEG-2 Sensitivity to Data Loss

In an MPEG-2 video stream, data loss reduces quality depending strongly on the type of the lost information. Losses in syntactic data, such as headers and system information, affect the quality differently than losses of semantic data such as pure video information (e.g., motion vectors, DCT coefficients, etc.). Furthermore, the quality reduction depends also on the location of the lost semantic data due, for instance, to the predictive structure of MPEG-2 video coded streams.

Figure 2 shows how network losses map onto visual information losses in different types of pictures. Indeed, 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, it 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 known as temporal propagation and is due to inter-frame predictions.

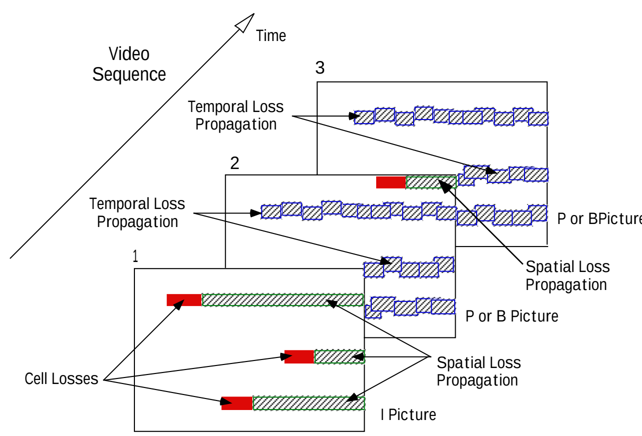

Figure 2: *Data loss propagation in MPEG-2 video streams.*

However, this error sensitivity may be dramatically reduced by means, for instance, of an adaptive syntactic information coding mechanism. Indeed, as stated before, slice headers and I-picture headers act as resynchronization points for, respectively, spatial and temporal propagations of errors.

3. PERCEPTUAL SYNTACTIC INFORMATION CODING

3.1. Preliminary Studies

In this work, we considered the adaptive insertion of slice headers only. In fact, the MPEG-2 standard allows to build slices with a variable number

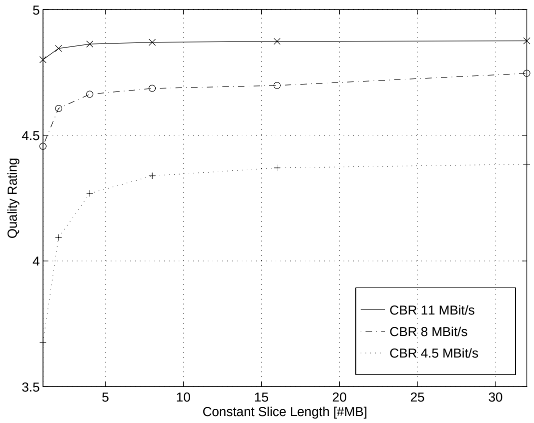

Figure 3: *Perceptual video quality versus number of macroblocks per slice for CBR MPEG-2 encoding. Simulations performed on the Ski sequence at 4.5, 8 and 11 MBits/s.*

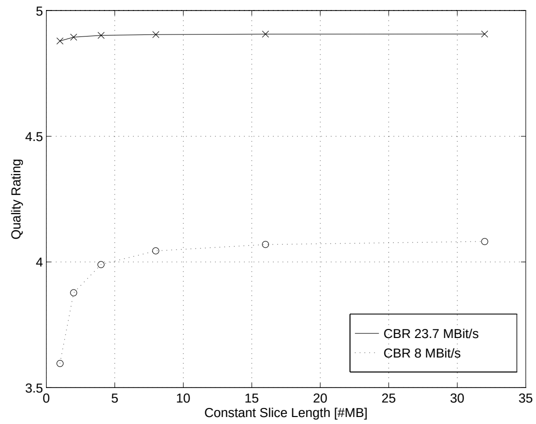

Figure 4: *Perceptual video quality versus number of macroblocks per slice for CBR MPEG-2 encoding. Simulations performed on the BasketBall sequence at 8 and 24 MBits/s.*

of macroblocks. The only restriction is that a new slice shall start on every new line of macroblocks and that slices shall occur in the bitstream in the order in which they are encountered. However, the greater the number of slices, the bigger the overhead. Indeed, each new slice introduces a 5- to 6-byte length header and resets the differential coding of the DC values and motion vectors.

Therefore, we first studied how the subjective video quality behaves with the number of macroblocks per slice with and without data loss in the bitstream. Several studies have shown that a correct estimation of subjective video quality has to incorporate some modeling of the Human Visual System [8]. In this work, we used a computational quality metric for moving pictures, namely the MPQM metric, which proved to behave consistently with human judgements [4]. The resulting quality rating is a value between 1 and 5 according to the ITU-R 500.3 standard [9] (the higher the value, the better the quality).

In our simulations, two 128 frame-long sequences conforming to the ITU-T 601 format have been used (i.e., BasketBall and Ski). These sequences have been encoded, as interlaced video, with a structure of 12 images per GOP and 2 B-pictures between every reference picture, both in CBR (constant bit rate) and open-loop VBR (i.e., a constant quantizer scale value is used over the whole sequence) modes. For that purpose, the TM5 MPEG-2 software encoder has been used [10].

Figure 3 shows how the perceptual video quality behaves according to the constant number of macroblocks per slice for CBR MPEG-2 encoding of the Ski and BasketBall sequences at several encoding bit rates. As stated before, we see that the quality is increasing with the slice length (i.e., the overhead decreases). Moreover, the impact of the slice length on the quality is greater for moderate bit rates (i.e., bit rates of interest).

However, for VBR encoding, the video quality is not influenced by the slice length since the bit rate is not fixed. Therefore, Fig. 5 shows how the mean encoding bit rate varies according to the slice length for a given quantizer scale factor.

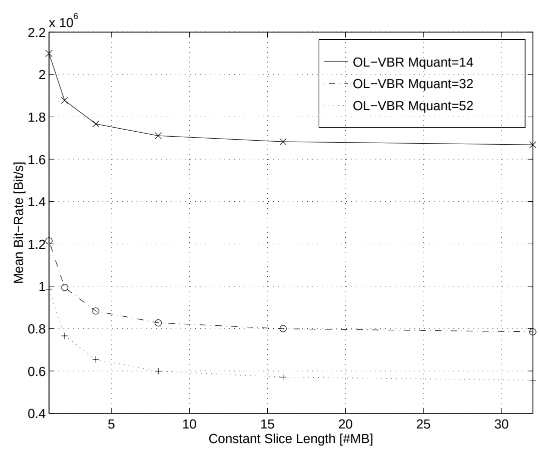

Figure 5: *Perceptual video quality versus number of macroblocks per slice for VBR MPEG-2 encoding at several quantizer scale values (14, 32 and 52). Simulations performed on the Ski sequence.*

In both cases, it may be noticed that the slice length influence saturates very quickly (around 10 macroblocks per slice).

Up to this point, we did not consider any data loss in the bitstream yet. However, as stated before, the bitstream robustness depends directly on the number of macroblocks per slice. Therefore, we considered the practical case of CBR MPEG-2 transmission over ATM networks [11]. The simulation framework is composed of two multimedia workstations and one ATM switch. The switching module is implemented as a multiplexer with limited buffer size. To generate cell losses in the multimedia streams, we load the multiplexing stage with background traffic provided by several On-Off sources. This type of source model is widely used to simulate a multiplex of traffic such as the one that could be found at the entrance of an ATM switch. Moreover, two state Markov source models encompass the peak cell rate parameter which is currently the most important traffic contract parameter [12].

Figure 6 presents the number of corrupted macroblocks as a function of the slice length for three different ATM cell loss ratios (CLRs). It is shown that the higher the CLR, the greater the impact of reducing the slice length on the number of corrupted macroblocks.

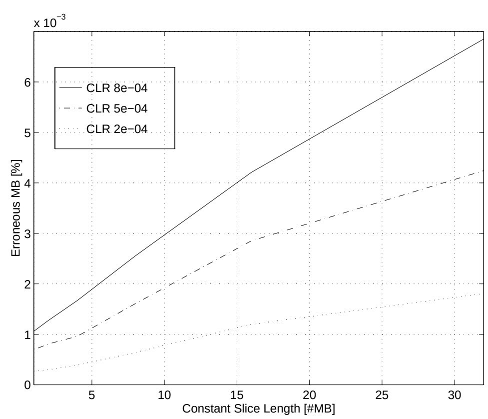

Figure 6: *Number of corrupted macroblocks versus slice length. Simulations performed on the Ski sequence.*

3.2. PSIC Mechanism

From the previous results, it is obvious that there is a tradeoff between the encoding quality and robustness in the case of CBR encoding. This tradeoff is illustrated in Fig. 7. Indeed, for a given encoding bit rate (5 Mbit/s.) and an imposed CLR ($5e^{-4}$), the perceptual video quality exhibits an optimum value corresponding to a constant slice length of four macroblocks. However, we see that this optimum changes with the CLR and the encoding bit rate.

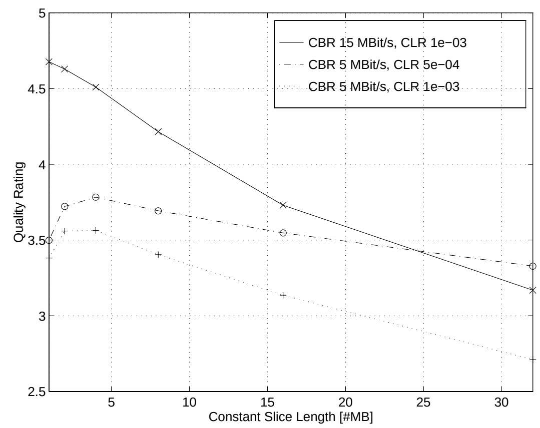

Figure 7: *Video quality versus the slice length for three different couples (constant encoding bit rate, ATM cell loss ratio). Simulations performed on the Ski sequence.*

Therefore, we designed a new mechanism, namely the Perceptual Syntactic Information Coding (PSIC), which adaptively modulates the number of macroblocks per slice in order to optimize the global perceptual video quality. This modulation is performed in the following way: for a given error concealment technique, as soon as the macroblock-based quality degradation (due to hypothetic macroblocks loss) reaches a given threshold, a new slice header is inserted. The quality degradation is measured by means of a macroblock-based average luminance value which, in a coarse approximation, corresponds to the simplest metric correlated with human perception (under the assumption that the viewer stands far enough from the monitor).

In our simulations, we considered the simplest temporal predictive concealment technique, proposed in the MPEG-2 standard, which consists in replacing a lost macroblock with the macroblock at the same location in the previous reference picture. The algorithm applied to this error concealment technique is summarized by the following equations.

A new slice is inserted as soon as:

$$\sum_{M_i \in S} \Delta_{M_i, M'_i} \geq Threshold, \qquad (1)$$

where, M_i is the current macroblock belonging to slice S, M'_i is the corresponding macroblock in the previous reference picture and,

$$\Delta_{M_i, M'_i} = \frac{1}{256} \sum_{p=1}^{256} M_i^p - \frac{1}{256} \sum_{p=1}^{256} {M'}_i^p. \qquad (2)$$

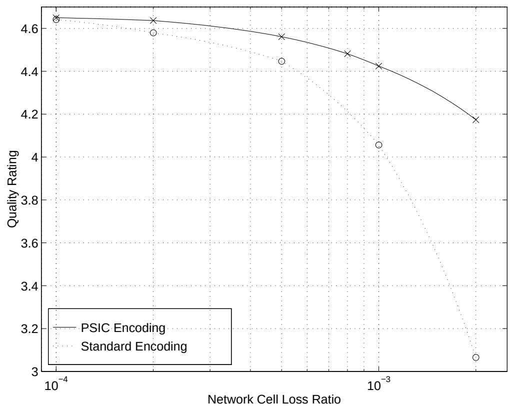

Figure 8: *Perceptual video quality versus the ATM cell loss ratio for both PSIC and fixed-length slice encoding at a CBR of 8 Mbit/s. Simulations performed on the Ski sequence.*

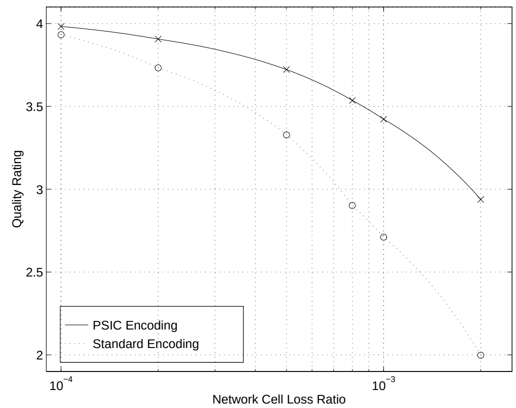

Figure 9: *Perceptual video quality versus the ATM cell loss ratio for both PSIC and fixed-length slice encoding at a CBR of 8 Mbit/s. Simulations performed on the BasketBall sequence.*

where, the exponent p represents the pixel position in the corresponding macroblock.

Furthermore, this mechanism takes the packetization process into account. Indeed, there is no need to have more than one slice header in the same network loss entity (e.g., AAL5-PDUs for ATM networks).

3.3. Performance Analysis

Figures 8 and 9 show the performance of the PSIC technique compared to the standard fixed-length slice CBR encoding process (in which a slice corresponds to a complete line of macroblocks) as a function of the network cell loss ratio. For that purpose, the same previously stated error concealment technique has been used in both cases.

One may notice that the proposed mechanism (PSIC) automatically adapts to the scene complexity and the picture type for a given encoding bit rate, an expected packet loss ratio and a maximum video degradation due to data loss in the bitstream. Moreover, the mechanism is flexible enough to still work properly with a different error concealment technique and a different macroblock-based quality degradation metric.

4. CONCLUSION AND FUTURE WORKS

In this paper, we presented a new flexible scene-complexity adaptive mechanism, namely the Perceptual Syntactic Information Coding (PSIC) mechanism.

We first studied how the subjective video quality behaves with the number of macroblocks per slice with and without data loss in the bitstream. We then introduced the proposed encoding mechanism. The PSIC adaptively modulates the number of macroblocks per slice in order to optimize the global video quality which depends on both the video encoding quality and the video degradations due to data loss during the transmission. The resulting mechanism appeared to behave consistenly for two TV-resolution sequences (BasketBall and Ski).

Further works are being carried out. We are working on a new macroblock-based video quality degradation measure (the one used in this paper is only valid in a coarse approximation). We are also investigating the PSIC's behavior in the case of VBR transmission over ATM networks. Finally, different error concealment techniques will be studied.

5. REFERENCES

[1] O.A. Aho and J. Juhola, "Error resilience techniques for mpeg-2 compressed video signal," in *IEE International Broadcasting Convention (IBC)*, pp. 327–332, January 1994.

[2] Iain E. G. Richardson and Martyn J. Riley, "Varying slice size to improve error tolerance of mpeg video," in *SPIE - Electronic Imaging Conference*, vol. 2668, January 1996.

[3] Iain E. G. Richardson and Martyn J. Riley, "Controlling the rate of mpeg video by

dynamic variation of sequence structure," in *International Picture Coding Symposium*, March 1996.

[4] C. J. van den Branden Lambrecht and O. Verscheure, "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System," in *Proceedings of the SPIE*, vol. 2668, (San Jose, CA), pp. 450–461, January 28 - February 2 1996. available on `http://tcomwww.epfl.ch/~verscheu/`.

[5] I. J. 1, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1, 2 and 3.* ISO/IEC JTC 1, 1994.

[6] Barry G. Haskell, Atul Puri and Arun N. Netravali, *Digital Video: an Introduction to MPEG-2.* Digital Multimedia Standards Series, Chapman and Hall, 1997.

[7] I. J. 1, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video.* ISO/IEC JTC 1, 1994.

[8] S. Comes, *Les traitements perceptifs d'images numerisées.* PhD thesis, Université Catholique de Louvain, 1995.

[9] ITU-R Commity, "ITU-R 500.3 Recommendations and Reports," in *ITU*, vol. XI, 1986.

[10] Chadd Fogg, "mpeg2encode/mpeg2decode version 1.1. available via anonymous ftp at ftp.netcom.com," in *MPEG Software Simulation Group*, June 1994.

[11] Xavier Garcia Adanez, Olivier Verscheure and Jean-Pierre Hubaux, "New Network and ATM Adaptation Layers for Real-Time Multimedia Applications: A Performance Study Based on Psychophysics," in *COST-237 workshop*, (Barcelona, Spain), November 1996. available on `http://tcomwww.epfl.ch/~verscheu/`.

[12] R. Slosiar, *Performance Analysis Methods of ATM-Based Broadband Access Networks Using Stochastic Traffic Models.* PhD thesis, TCOM Laboratory, Swiss Federal Institute of Technology, Lausanne, 1995.