Published in NOSSDAV 98, pp. 257-264, Cambridge, July 1998

MPEG-2 Video Services over Packet Networks: Joint Effect of Encoding Rate and Data Loss on User-Oriented QoS

Olivier Verscheure, Pascal Frossard and Maher Hamdi
Institute for computer Communications and Applications
Swiss Federal Institute of Technology

Abstract

We address the problem of video quality prediction and control for high resolution video transmitted over lossy packet networks. We analyze how the user-perceived quality is related to the average encoding bitrate for VBR MPEG-2 video. We then show why simple distortion metrics (e.g., PSNR) may lead to inconsistent interpretations. Furthermore, for a given coder setup, we analyze the effect of packet loss on the user-level quality. We then demonstrate that, when jointly studying the impact of coding bit rate and packet loss, the reachable quality is upperbound and exhibits one optimal coding rate for a given packet loss ratio.

1 Introduction

The choice of the compression algorithm depends on the available bandwidth or storage capacity and the features required by the application. The MPEG-2 [1] standard [1], 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 [2]. However, much work remains to be done to optimize these audiovisual applications as the users expect an adequate audiovisual quality at the lowest possible cost. In the case of video transmission over packet networks, the User-oriented Quality of Service (U-QoS) results both from the video encoding quality and the degradations due to packet loss, delay and delay jitter during the transmission. The most economic offering can thus only be found by considering the entire system and not by optimization of individual system components in isolation [6].

Our work focuses on video quality prediction and control for high resolution packet video transmitted over lossy networks.

This paper is organized as follows: In Sec. 2, we introduce the MPEG-2 video and system standards. We briefly describe the impact of data loss on the reconstructed video sequence. Finally, useful video quality metrics are described, among them is the MPQM which is based on a vision model. The study of the impact of MPEG-2 rate and data loss on quality is the subject of Sec. 3. Section 4 deals with the joint impact analysis of both MPEG-2 rate and data loss on video quality. Concluding remarks are given in Sec. 5.

2 MPEG-2 over Packet Networks

2.1 MPEG-2 Backgrounder

An MPEG-2 video stream is 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. Motion estimation is performed on macroblocks while the DCT [2] is calculated on blocks. The resulting DCT coefficients are quantized and variable length coded. The quantizer comes from the multiplication of a Quantizer Scale, MQUANT, and the corresponding element of a Quantizer Matrix. In general, the higher the MQUANT value, the

[1] MPEG stands for Moving Picture Experts Group

[2] DCT stands for Discrete Cosine Transform

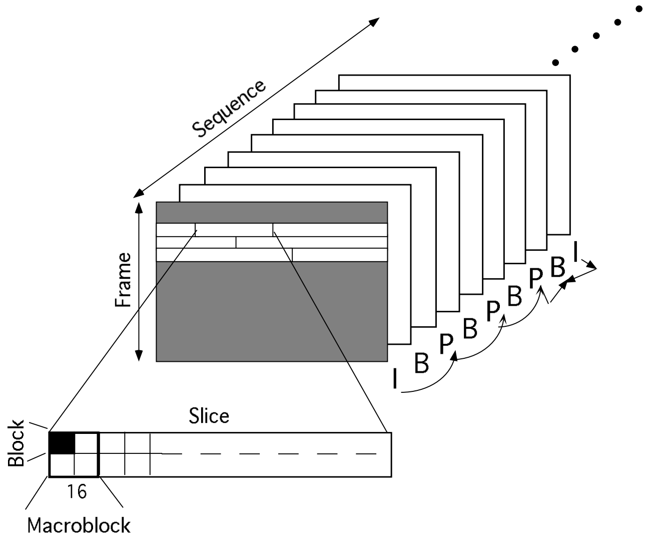

Figure 1. MPEG-2 video structure.

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 packets and then subdivided into fixed-length TS packets. 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. In general, when a header is damaged, the underlying information is lost.

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 of 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 on the location of the lost semantic data due, not only to the predictive structure of MPEG-2 video coded streams, but also to the visual relevance of the data.

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 known as temporal propagation and is due to inter-frame predictions.

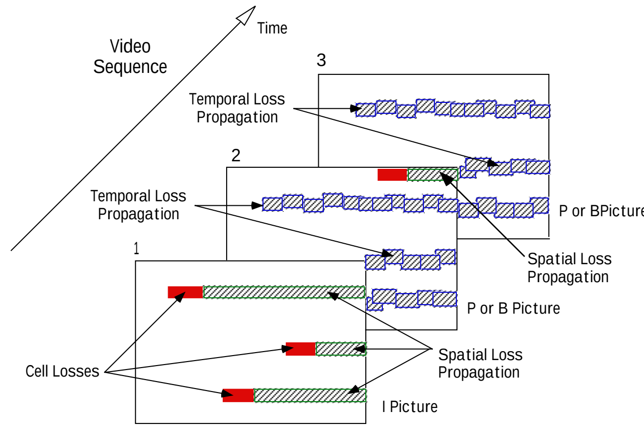

Figure 2. Data loss propagation in MPEG-2 video streams.

However, the error visibility may be dramatically reduced by means of error concealment techniques. These error concealment algorithms include, for example, spatial interpolation, temporal interpolation and early resynchronization techniques. The MPEG-2 standard proposes an elementary error concealment algorithm based on motion compensated techniques. Mainly, it estimates the motion vectors for the lost macroblock by using the motion vectors of neighbouring macroblocks in the affected picture (provided 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 [3].

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.

2.3 Video Quality Metrics

Traditionally, the quality metric used for audiovisual signals is the Peak Signal to Noise Ratio (PSNR). However, many works have shown that such a metric was poorly correlated with human perception. Indeed, the PSNR metric does not take the visual masking phenomenon into consideration. In other words, every single errored pixel contributes to the decrease of the PSNR, even if this error is not

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

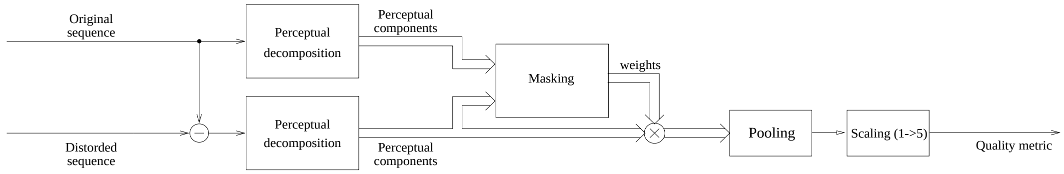

Figure 3. Moving Pictures Quality Metric (MPQM) block diagram

perceived. Hence, recent research has addressed the issue of video quality assessment by means of human correlated metrics.

Several studies have shown that a correct estimation of subjective quality has to incorporate some modeling of the Human Visual System [4]. A spatio-temporal model of human vision has been developed for the assessment of video coding quality [10, 11]. It was then used to build a computational quality metric for moving pictures [11] which proved to behave consistently with human judgments. Basically, the metric, termed Moving Pictures Quality Metric (MPQM), first decomposes an original sequence and a distorted version of it into perceptual channels. A channel-based distortion measure is then computed, accounting for contrast sensitivity and masking. Finally, the data is pooled over all the channels to compute the quality rating which is then scaled from 1 to 5 [8] (see Fig. 3). This quality scale is used for subjective testing in the engineering community (see Table 1).

Rating	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

Table 1. Quality scale that is generally used for subjective testing in the engineering community

3 Impact of MPEG-2 Rate and Data Loss on Quality

In this section, we first describe the experimental setup used throughout this work. We then study how the video quality behaves according to the quantizer scale factor (MQUANT) in an MPEG-2 OL-VBR [4] encoding scheme. We also analyze how the average encoding bit rate is affected by this MQUANT. We then derive a mathematical relation modeling the impact of the average variable rate on the video encoding quality. Finally, we study how the video quality decreases when the data loss ratio is increased, for a fixed average encoding bit rate.

[4]OL-VBR stands for Open-Loop Variable Bit Rate (constant quantizer scale over the whole sequence).

3.1 Experimental setup

The experimental testbed is composed of four parts (see Fig. 4):

- An MPEG-2 software encoder, which is composed of an open-loop VBR TM5 video encoder [3] and a transport stream encoder. Four 100 frame-long sequences conforming to the ITU-T 601 format were used (i.e., Football, News, Ski and Barcelona). All these sequences are very different in terms of spatial and temporal complexities. They were encoded, as interlaced video, with a structure of 12 images per GOP and 2 B-pictures between every reference picture in an OL-VBR mode. The following MQUANTs were used: 6, 10, 16, 20, 28, 32, 36, 40 and 48. Motion vectors were generated for all intra-coded macroblocks. It is to be noted that the OL-VBR encoding quality is not affected at all when introducing these extra motion vectors. Before being transmitted, each MPEG-2 video bitstream was encapsulated into 18800-bytes length Packetized Elementary Stream (PES) packets and divided into fixed length Transport Stream (TS) packets by the MPEG-2 system encoder.

- A model-based data loss generator was used to simulate packet network losses. For this purpose, we used a two-state Markovian model (Gilbert model [5], see Fig. 5).

  States 0 and 1 respectively correspond to a correct and an incorrect packet reception. The transition rates between the states control the lengths of the bursts of errors. Hence, there are three parameters to be controlled: the packet loss size (PLS), the packet loss ratio ($PLR = \frac{p}{p+q}$) and the average length of a burst of errors ($ABL = \frac{1}{q}$). In our simulations, we imposed

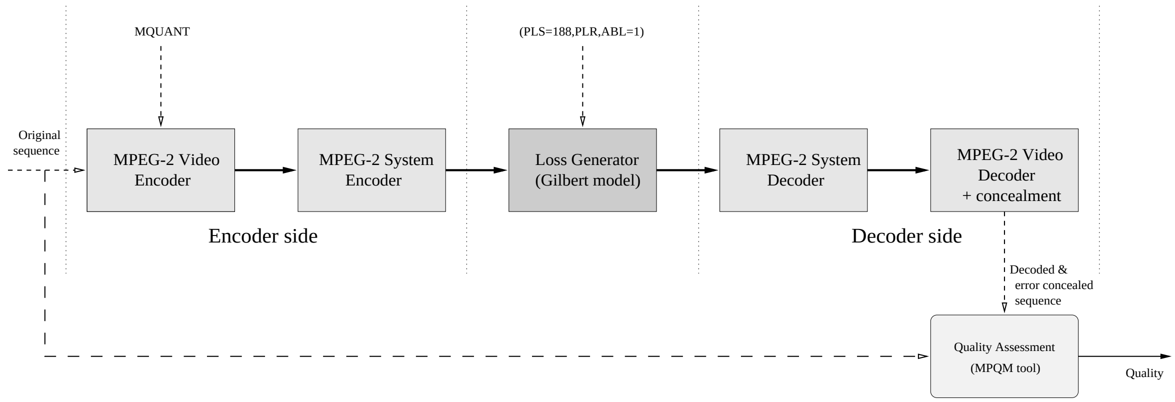

Figure 4. Experimental testbed.

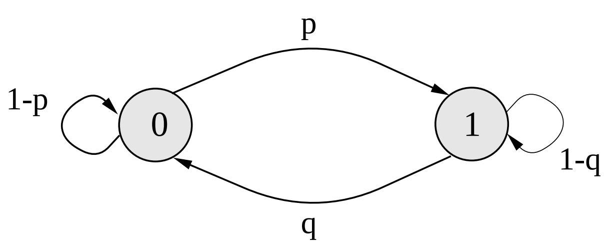

Figure 5. Two-State Makov Chain: Gilbert Model.

a non-bursty ($ABL = 1$) TS packets ($PLS = 188$ bytes) loss process and made the packet loss ratio vary between 10^{-2} and 10^{-7}.

- Video quality was evaluated by the MPQM tool (see Sec. 2.3).
- The last part is an MPEG-2 software decoder constituted by both a TS decoder and a video decoder. The video decoder provides the motion compensated concealment technique briefly explained in Sec 2.2.. This technique was chosen for different reasons. The first is to be consistent with real implementations. The second is to be able to perform the perceptual measurements. Indeed, the vision model currently developed and the derived metrics have been tested for errors below what is called the *suprathreshold* [5]. The problem is that, in general, the degradations due to data losses generate highly visible artefacts (i.e., holes) in the sequence and these errors are all above this suprathreshold. By using concealment techniques, most of the artifacts may be considered as being below the suprathreshold of vision, making the perceptual measure accurate.

[5]Two to three times above the threshold of vision which corresponds to the threshold of visibility of the noise

3.2 MPEG-2 VBR Encoding Impact on Video Quality

First, we study how the OL-VBR encoding process influences video quality. Figures 6 and 7 show how the quality is affected by the MQUANT parameter using, respectively, the PSNR metric and MPQM tool to measure it. While the PSNR versus MQUANT curve may be represented by a decreasing exponential [9], it is to be noted that the MPQM metric exhibits a linear relationship with MQUANT. We verified this important behaviour for the four sequences constituting our testbed. The same characteristic has recently been observed through user's subjective evaluation [7]. Computer simulation results as well as the corresponding fits are represented on Fig. 7 for both the "Barcelona" and "News" sequences. The encoding quality is approximated by a function of the form:

$$Q_E = \chi_Q \times MQUANT + Q_0 \quad (1)$$

where parameters χ_Q and Q_0 have been obtained by minimizing the mean square error (see Fig. 7). The slope χ_Q is directly related to the complexity of the sequence: the higher the encoding complexity, the higher the absolute value of χ_Q. This remark may be verified on the graph. The video sequence "News" is a *Head and Shoulder* type of sequence and does not contain any high spatio-temporal complexities. The absolute value of the slope is thus smaller than for the "Barcelona" video sequence. The value of Q_0 will always be close to 5 (highest quality).

This linear relation between the video quality and the quantizer scale factor may have several impacts on the design of, for instance, perceptual rate controllers or consistent quality regulators.

Now, we have an idea of how the encoding quality behaves according to the MQUANT. We need then to study how the average output bit rate if affected by this

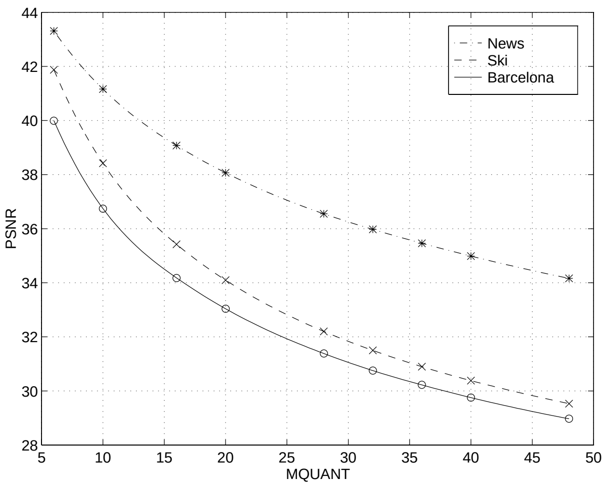

Figure 6. PSNR versus quantizer scale factor for 3 different scenes.

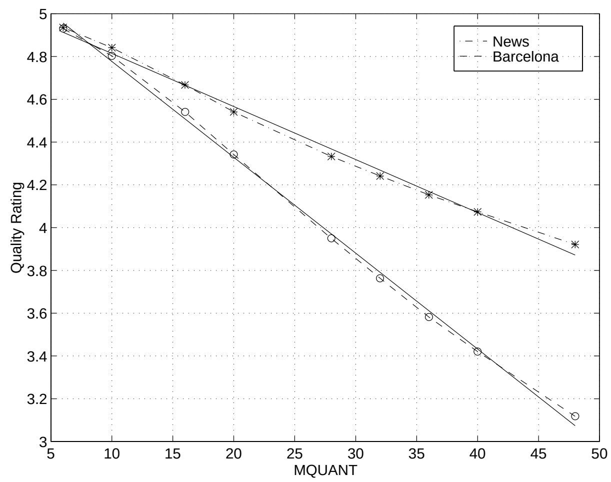

Figure 7. MPQM versus quantizer scale factor for 2 different scenes. Fitting parameters for Eq.1: News: (χ_Q=-0.025, Q_0=5.062), Barcelona: (χ_Q=-0.045, Q_0=5.226)

MQUANT. In [9], it was demonstrated that a power function curve was a good approximation to represent the relation between the quantizer scale factor and the average bit rate:

$$\bar{R} = \chi_R \times MQUANT^{-\xi_R} \tag{2}$$

in which $\bar{R}$ represents the average output bit rate and the parameters χ_R and ξ_R are related to the encoding complexity of the scene.

Figure 8 illustrates this behaviour very well. The parameters χ_R and ξ_R have been obtained by minimizing the mean square error.

Finally, by combining equations (1) and (2), we derive a model for describing how the video quality behaves according to the average encoding bit rate:

$$Q_E = \chi_Q \times \left(\frac{\bar{R}}{\chi_R}\right)^{-\frac{1}{\xi_R}} + Q_0 \tag{3}$$

As stated before, the three main parameters χ_Q, χ_R and ξ_R are somehow related to the spatio-temporal complexity of the sequence (Q_0 will always be around 5.0). However, in this work, we did not investigate this relation any further. This work would not be trivial, as it involves not only encoding complexities but also visual masking phenomenons. We are currently investigating such an extension.

Computer simulation results and the corresponding fitting curve using the equation herebefore are represented in Fig. 9.

An important result that can be extracted from the graph is that the perceptual quality saturates at high bit rates. Increasing the bit rate may thus result, at some point, in a waste of bandwidth since the end-user does not perceive an improvement in quality after a certain bit rate. However, such saturation of quality is not well captured by the PSNR.

3.3 Data Loss Impact on Video Quality

Up to this point, we did not consider any data loss in the video stream. Figure 10 illustrates how the video quality is affected by uniformly distributing TS packet losses over MPEG-2 transport streams. It is shown that, on a semi-logarithmic scale and for a given MQUANT (average bit rate), first the video quality remains constant with the PLR. This constant value corresponds to the encoding quality. Then, beyond a certain PLR, the perceptual quality quickly drops.

The higher the MQUANT value, the higher the PLR after which the quality drops. Indeed, since the PLR is equal to the ratio between the packet loss rate (number of lost packets per second) and the packet rate, then the lower the bit rate (the higher the MQUANT), the lower the packet loss rate for a fixed PLR. Also, the lower the packet loss rate, the lower the number of lost packets per frame on average. Therefore, the higher the MQUANT, the higher the PLR for an equivalent perceived degradation (see Fig. 10).

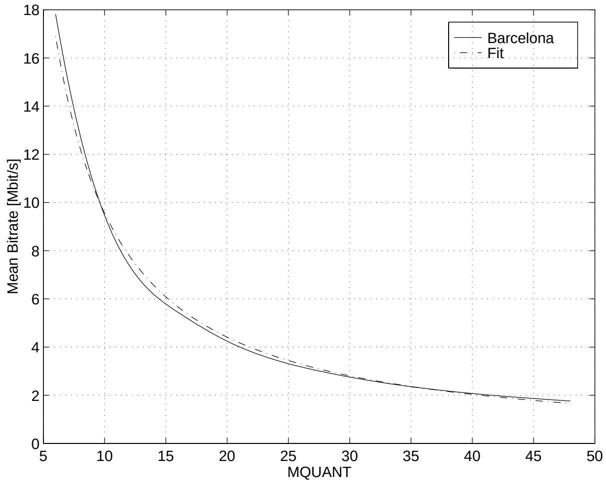

Figure 8. Average output encoding bit rate versus quantizer scale factor (MQUANT) for Barcelona. Fitting parameters for Eq.2: ($\chi_{\mathbf{R}}$=124.762, $\xi_{\mathbf{R}}$=1.116)

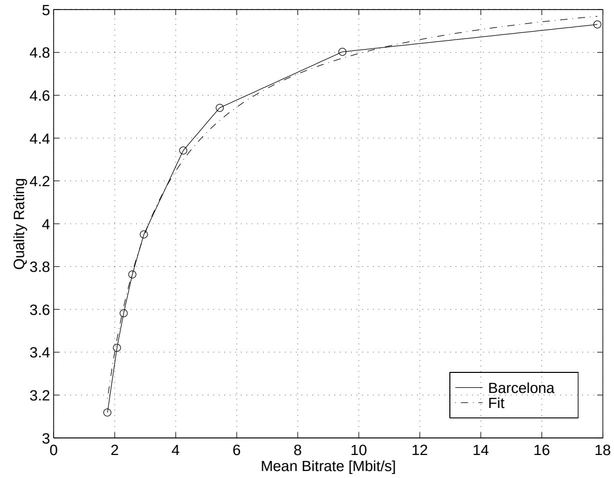

Figure 9. MPQM video quality versus average output encoding bit rate for the Barcelona sequence. Fitting parameters for Eq. 3: ($\chi_{\mathbf{Q}}$=-0.045, $\mathbf{Q_0}$=5.225, $\chi_{\mathbf{R}}$=124.761, $\xi_{\mathbf{R}}$=1.116)

Hence, the relation between video quality and PLR may be represented by a straight line on a linear scale:

$$Q = Q_E + \chi_L^* \times PLR, \tag{4}$$

where Q_E corresponds to the encoding quality (given by Eq. 3) and χ_L^* depends on both the complexity of the sequence and the average bit rate. In other words, for a given sequence and a fixed MQUANT, the video quality, averaged over the whole sequence, linearly decreases with the PLR.

This relation still holds if we multiply the PLR by a constant. We observed that, for a given $MQUANT$, the relation between the end-to-end video quality, Q, and the product $\bar{R} \times PLR$ could be well approximated by a straight line of slope χ_L. Therefore, Eq. 4 becomes:

$$Q = Q_E + \chi_L \times (\bar{R} \times PLR), \tag{5}$$

where χ_L is almost independent of the $MQUANT$ especially for low to medium bit rates.

4 Joint Impact of MPEG-2 Rate and Data Loss on Quality

In this section, we demonstrate why a joint analysis of the impact of both the MPEG-2 encoding bit rate and the data loss ratio on the video quality is the only way to get meaningful conclusions. We explain, for example, why the video quality may decrease when the encoding bit rate is increased in an error-prone environement.

4.1 Joint Impact Analysis

As stated at the end of the previous section, the PLR and the encoding bit rate (packet rate) are intimately related to each other in regards to their impact on video quality. For example, the higher the bit rate, the higher the encoding quality (up to saturation) but the lower the PLR after which the video quality quickly drops, and conversely. Therefore, the relation between quality and encoding bit rate for a non-zero PLR should somehow exhibit an optimum value. This behaviour is illustrated on Fig. 11. We indeed see that the video quality first increases (encoding quality) with the average bit rate and then decreases after around 4 MBits/s for the "Barcelona sequence" (data loss). This optimal average bit rate directly depends on the sequence type. We observed that it was fairly independent of the PLR though.

Such a result is crucial for the design of network-aware rate controllers, efficient error concealment algorithms, etc.

4.2 Tri-Dimensional Representation

The purpose of this subsection is to put all the results together and represent them by a single graph. Thus, by putting together Eq. 3 and Eq. 5, we obtain a model of the end-to-end video quality Q:

$$Q = \chi_Q \times \left(\frac{\bar{R}}{\chi_R}\right)^{-\frac{1}{\xi_R}} + Q_0 + \chi_L \times (\bar{R} \times PLR), \tag{6}$$

in which the two first terms of the sum represent the encoding quality (see Eq. 3).

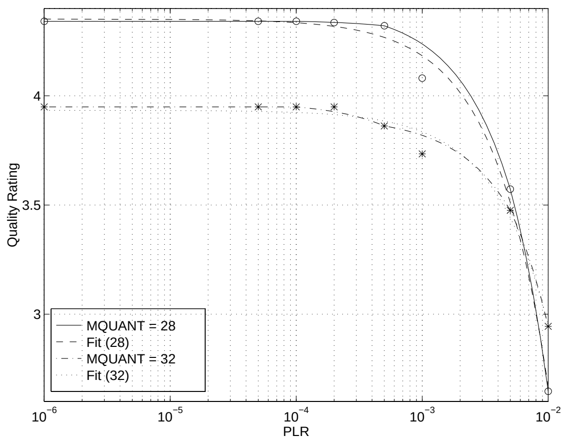

Figure 10. MPQM versus PLR (ALB=1, PLS=188) for MQUANTs={28, 32} using the Barcelona sequence. Fitting parameters for Eq.4: MQUANT=28, ($\mathrm{Q_E}$=4.352, χ^*_{L}=-168.162) and for MQUANT=32, ($\mathrm{Q_E}$=3.934, χ^*_{L}=-98.351). Sequence: Barcelona

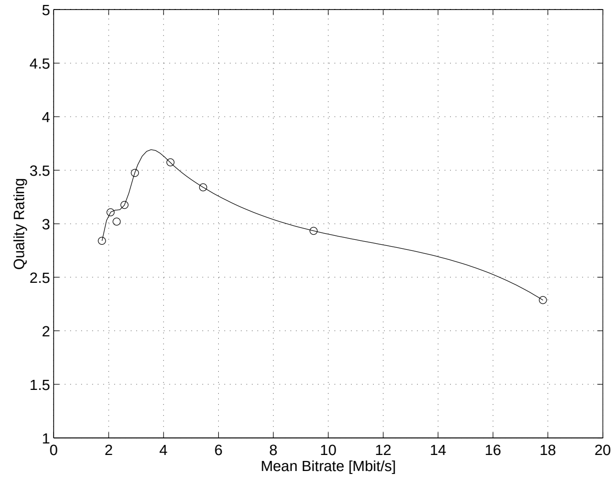

Figure 11. MPQM versus average encoding bit rate for $\mathbf{PLR = 5 \times 10^{-3}}$ for the Barcelona sequence.

We then performed a complete set of measurements in order to verify this relation. The same simulation setup as the one presented in the previous section has been used. Figure 12 presents the resulting surface for the "Barcelona" sequence while Fig. 13 shows its corresponding fit using relation (6).

Several results may be extracted from these graphs. Most of these results have already been discussed throughout this paper. In general, when considering video transmission over lossy networks, not only it is bandwidth consuming to increase the encoding bit rate above a certain threshold due to saturation of quality (which varies according to the scene complexity), it may also be quality consuming. In other words, when the user-oriented QoS is not high enough, an increase of the encoding bit rate at a fixed PLR may even degrade the quality, depending on the position of the working point on the 3D graph presented herebefore. There is an optimal bit rate to be determined that maximizes the end-user perception of the service under certain given network conditions (i.e., network impairments).

Such a conclusion is general enough to be applied to a different encoding system.

5 Conclusion and Future Works

The combined effect of the coding bitrate and the network impairements on the user-perceived quality is still not well understood. However these results are needed for the design and deployment of packet video services. One of the common misleading intuitions is that increasing the coder bit rate enhances image quality. In this paper we have shown that this intuition is proper to a lossless communication channel and that the quality-rate function is no longer a strictly increasing function when video packets are subject to loss.

The major conclusion is that image quality cannot be improved by acting on the coding bit rate only: increasing the bit rate above a certain threshold results in quality degradations. For a given packet loss ratio, there is a quality-optimal coding rate that has to be found. Although the relationship between coding bit rate, packet loss ratio and user-level quality is intrinsically complex, it can be characterized by a simple expression and parameters set. These parameters seem to depend on global properties of the video sequence (e.g., spatio-temporal complexities). Such parameters have to be predicted when video is coded and transmitted in real-time over lossy networks. Therefore, this work is being extended to on-line prediction of the 3D quality graph in the context of MPEG-2, as well as other emerging encoding standards.

References

[1] I. J. 1. *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1, 2 and 3.* ISO/IEC JTC 1, 1996.

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

[3] Chadd Fogg. mpeg2encode/mpeg2decode. In *MPEG Software Simulation Group*, 1996.

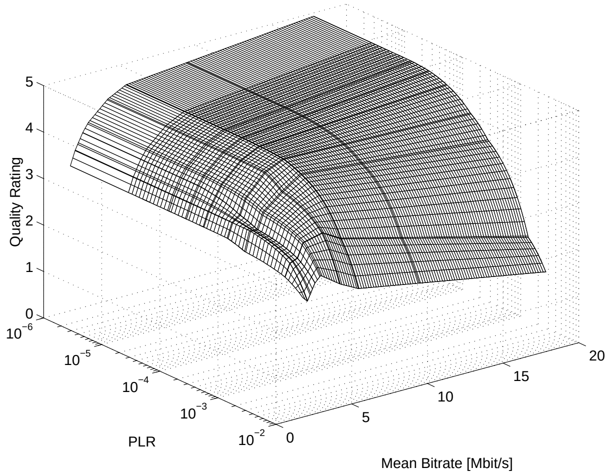

Figure 12. Q versus average bit rate and PLR: Simulations on the Barcelona sequence.

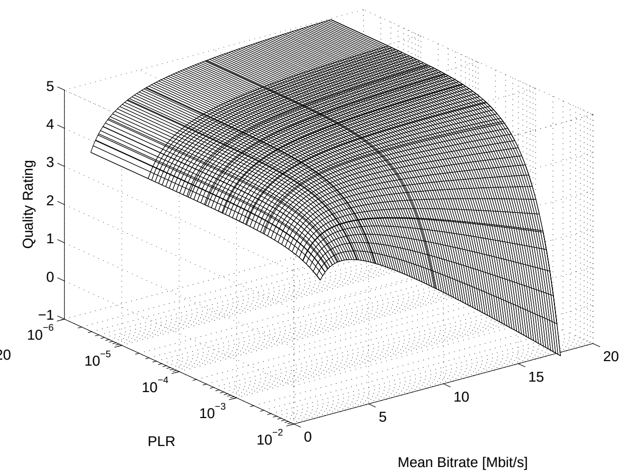

Figure 13. Q versus average bit rate and PLR: Fitting function with parameters: ($\chi_{\mathbf{Q}}$=-0.045, $\mathbf{Q_0}$=5.22, $\chi_{\mathbf{R}}$=124.76, $\xi_{\mathbf{R}}$=1.12 and $\chi_{\mathbf{L}}$=-33.9).

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

[5] J. W. Roberts, J. Guibert and A. Simonian. Network Performance Considerations in the Design of a VBR Codec. In *Queuing Performance and Control in ATM*, pages 77–82. J. W. Cohen and C. D. Pack, June 1991.

[6] G. Karlsson. Asynchronous Transfer of Video. *IEEE Communications Magazine*, 34(8):118–126, August 1996.

[7] Kentarou Fukuda, Naoki Wakamiya, Masayuki Murata and Hideo Miyahara. On Flow Aggregation for Multicast Video Transport. In *6th IFIP International Workshop on Quality of Service (IWQoS)*, May 1998.

[8] M. Ardito and M. Barbero and M. Stroppiana and M. Visca. Compression and Quality. In *Proceedings of the International Workshop on HDTV 94*, Brisbane, Australia, October 1994.

[9] Shigeyuki Sakazawa, Yasuhiro Takishima, Masahiro Wada and Yoshinori Hatori. Coding Control Scheme for a Multi-Encoder System. In *7th International Workshop on Packet Video*, pages 83–88, March 1996.

[10] C. J. van den Branden Lambrecht. A Working Spatio-Temporal Model of the Human Visual System for Image Restoration and Quality Assessment Applications. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 2293–2296, Atlanta, GA, May 7-10 1996.

[11] 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*, volume 2668, pages 450–461, San Jose, CA, January 28 - February 2 1996.