

Modelling H.264/AVC sensitivity for error protection in wireless transmissions

Cyril Bergeron and Catherine Lamy-Bergot

THALES Land and Joint Systems, EDS/SPM, F-92704 Colombes Cedex.

Email: {cyril.bergeron,catherine.lamy}@fr.thalesgroup.com.

Abstract—A simple method is proposed to optimise the protection levels on the different parts of an H.264/AVC bitstream for transmission over an error-prone channel. This method relies on a semi-analytical model of the video stream sensitivity, that allows to predict the resulting distortion depending on the channel errors viewed by the video decoder. Valid for a frame, a partition or a Group of Pictures (GOP) by taking into account the dependencies existing between partitions and frames, the model allows to set the adapted protection for each partition or frame to minimise the overall sequence distortion for a given erroneous context by joint source channel optimisation of the compression and protection mechanisms.

The method is then applied to determine the best trade-off of channel and source coding rates for a given operating point or protection rate, with equal or unequal error protection.

I. INTRODUCTION

The transmission of multimedia data over bandwidth-limited and error-prone channels has imposed to reconsider Shannon's separation principle [1] that recommended independent design of the source coding (compression) and channel coding (protection) operations. Still, to allow for compatibility with existing standards, and deployment on existing architectures where network layers can be present between the source and channel coders, an integration into a unique joint coder is not considered here, and the compression and protection are kept apart, albeit in cooperation. Joint source and channel coding ensures that the impact of errors, almost unavoidable in wireless channels, is taken into account by combining efficiently compression and protection from the rendering point of view. As a matter of fact, the classical source rate control algorithms proposed in the absence of transmission errors such as [2] rely on the assumption that Forward Error Correction (FEC) tools let the packets arrive error free at the video decoder. While of particular interest for wired transmissions and broadcasting, these solutions do not take into account either the potentially important distortion effects introduced by a residual bit error probability unavoidable in low-bandwidth transmissions or the different sensitivities of the bitstream.

A first direction for joint tandem coding, following the idea that video decoders suffer mostly from packet losses [3], proposes source rate control solutions in absence of transmission errors, or establishes packet dropping mechanisms [4][5]. However, this network-oriented approach does not allow to take advantage of the most recent transport protocols such as UDPLite or DCCP which allow erroneous payloads to reach the application level where robust decoders can exploit them.

A second family of joint tandem coding schemes, to which this paper belongs, relies on the use of FEC tools to ensure that the bit and/or packet error probabilities viewed by the video decoder are below a given threshold. The most efficient schemes choose channel coding rates based on the analysis of the bitstream sensitivity, the key problem being the evaluation of this sensitivity representing the joint source and channel distortion due to the transmission of compressed video over an erroneous channel. Previous work, following general DCT-based approaches [6][7] or for given predictive standards [8][9] propose a sensitivity definition based on an analytical formulation for each frame or with a water-filling approach. However these solutions, eventually due to their generic non-standard based approach, either require experimentations to fit the model, or do need specific information (e.g. intra refresh rate) to fully take into account the different dependencies existing in the bitstream.

In this paper, a simple semi-analytical model predicting the distortion in a reconstructed video by deriving the impact of errors on the different partitions/frames of the H.264/AVC standard, based on their respective sensitivity to errors, and the extension to the distortion prediction of a GOP are proposed in Section II. When used with FEC protection, the formulas allow to specify the protection allocation minimising the GOP or video sequence distortion, by application of the protection rate adapted to the different sensitivity levels, as presented in Section III with simulation results obtained over an Additive White Gaussian Noise (AWGN) channel with Rate-Compatible Punctured Convolutional (RCPC) codes. Finally, some conclusions are drawn in Section IV.

II. SENSITIVITY FORMULATION

We propose to estimate the average expected end-to-end distortion \hat{D}_{S+C} after the source and channel coding operations for a video sequence. For sake of simplicity, each frame is assumed coded into a single slice (or Network Abstraction Layer (NAL) in the H.264/AVC standard), even though the results can be extended to the multiple slices case, as done with Data Partitioning.

The distortion \hat{D}_{S+C} for a frame (or NAL) transmitted over an error prone channel can be derived by taking into count the different distortion D_i corresponding to respective associated error event probability P_i .

$$\hat{D}_{S+C} = \sum_{i \in \mathbb{N}} D_i \cdot P_i$$

Theoretically, each single bit error, and their different combinations is a different error event, whose impact on the resulting decoded picture (with or without error concealment) should be taken into account. This discrimination level being far too complex to be modelled, it is proposed to assume that the errors can be grouped and averaged, considering the distortion resulting from errors in the frame, whether for leading to a loss of the NAL with D_{loss} , or to partial corruption of the NAL with D_{corr} , and the distortion inherent to compression operation, impacting even correctly received NALs with D_o . For P_c (resp. P_l) the probability to receive correctly (resp. to lose completely) a NAL, the following joint source and channel distortion, or *sensitivity* is obtained as:

$$\hat{D}_{S+C} = P_c.D_o + P_l.D_{loss} + (1 - P_c - P_l).D_{corr} \quad (1)$$

The resulting distortion is then expressed in terms of Mean Squared Error (MSE) or Peak Signal to Noise Ratio (PSNR):

$$\hat{MSE} = \sum_{i=1}^M \sum_{j=1}^Q \frac{(pl^*(i,j) - pl(i,j))^2}{M \times Q}$$

$$PSNR = 10 \log_{10} \left(\frac{255^2}{\hat{MSE}} \right)$$

with M, Q the width and height of the video frame, and $pl(i, j)$ (pl^*) the luminance of original (reconstructed) frames pixels.

Let now express the probabilities depending on the transmission channel. Considering as an example a memoryless erroneous channel with bit error event probability P_e such as the Binary Symmetric Channel (BSC) or AWGN channel, we have:

$$P_c = (1 - P_e)^n$$

where n is the frame size in bits and $P_e = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_S}{N_0}} \right)$ for a Signal-Noise Ratio $SNR = E_S/N_0$ without channel coding.

The probability P_l to lose a NAL is expressed by following the observations made in [11], where it was found that H.264/AVC Intra and Predicted frames could be partially noised (fraction p of the frame) without desynchronization of the bitstream, resulting only in visual errors (artifacts) in the reconstructed image: it is assumed that frames with more errors than the fraction p are lost due to desynchronization, while frames with less errors are corrupted, which gives:

$$P_l = 1 - (1 - P_e)^{(1-p)n},$$

leading to the sensitivity expression:

$$\hat{D}_{S+C} = (1 - P_e)^n D_o + (1 - (1 - P_e)^{(1-p)n}).D_{loss} + ((1 - P_e)^{(1-p)n} - (1 - P_e)^n).D_{corr} \quad (2)$$

A. Intra (I) and Predicted (P) frames

Taking into account the empirical observation that $MSE_{corr} \simeq MSE_o$ for H.264/AVC Intra and Predicted frames, and the estimation from [11] of fraction p for Intra

frames as $1 - \beta_0 \simeq 0.25$ and as $1 - \beta_i \simeq 0.15$ for i^{th} Predicted frames P_i , we can express the Intra frame sensitivity \hat{D}_{Intra} :

$$\hat{D}_{Intra} = (1 - P_e)^{\beta_0 n}.D_o + (1 - (1 - P_e)^{\beta_0 n}).D_{loss} \quad (3)$$

Similarly, the sensitivity expression of the i^{th} Predicted frame P_i of a GOP, when the previous ones are correct, is easily obtained as:

$$\hat{D}_{P_i} = (1 - P_e)^{\beta_i n_i}.D_{o_i} + (1 - (1 - P_e)^{\beta_i n_i}).D_{loss_i} \quad (4)$$

with size n_i the size of the i^{th} P-frame, D_{o_i} (resp. D_{loss_i}) the distortion observed when the frame is correct (resp. lost) with the previous ones are correct.

The sensitivity of a H.264/AVC encoded frame is hence derived with solely estimating the obtained distortion for best (no transmission error) and worst (frame lost) transmission conditions, and the frame length. As illustrated in Fig. 1 for the MSE sensitivity of sequence 'Foreman' first Intra frame in QCIF format for different quantization parameters (QP), our analytical expression \hat{D}_{Intra} is very close to experimental simulation results obtained with JM 10.1 [10] for 6000 trials. Similarly, the average MSE sensitivity obtained for the sequence 'Foreman' first GOP 14 P frames in QCIF format for $QP_I = 28, QP_P = 30$ over 6000 trials is very close to the analytical expression averaged over the same 14 frames, as illustrated in Fig. (2). Note that the higher value of D_{loss_i} for P frames is due to a better error concealment thanks to the GOP Intra frame being correct when deriving \hat{D}_{P_i} .

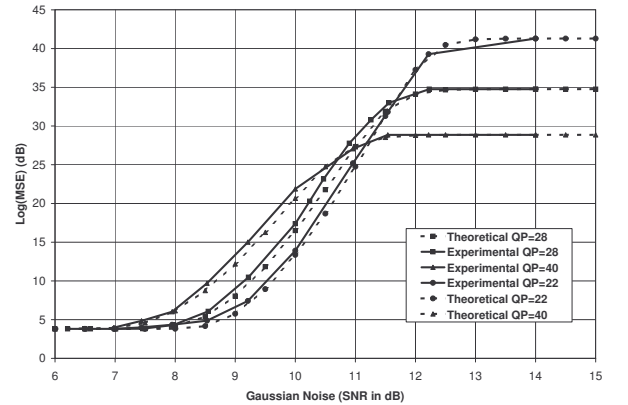


Fig. 1. Intra frame MSE sensitivity, First 'Foreman' sequence frame.

B. Group of frames

Let us now derive the sensitivity for a GOP made of an Intra frame followed by N Predicted (P) frames, or more generally a group of frames. In practice, P-frames and their sensitivity do depend on the previous frames: should one P-frame be badly received, the following ones, eventhough correctly transmitted, will not be perfectly reconstructed. As a consequence, we propose to assume here that if a frame is lost, the distortion contribution of eventual following frames is negligible. The impact of previous frames being incorrectly received are consequently taken into account by use of conditional probability on the previous ones to be correct.

With help of Eq. (3) and (4) with $\beta_0 = \alpha$, the distortion for a GOP is expressed as:

$$\begin{aligned} \hat{D} &= P_c^{(\beta_0)} \cdot D_{o_o} + (1 - P_c^{(\beta_0)}) \cdot D_{loss_0} \\ &= P_c^{(\beta_0)} [P_c^{(\beta_1)} D_{o_1} + (1 - P_c^{(\beta_1)}) D_{loss_1}] + (1 - P_c^{(\beta_0)}) D_{loss_0} \\ &= \dots \\ &= \left(\prod_{i=0}^N P_c^{(\beta_i)} \right) D_{o_N} + \sum_{i=0}^N \left[\prod_{j=0}^{i-1} P_c^{(\beta_j)} (1 - P_c^{(\beta_i)}) D_{loss_i} \right] \end{aligned} \quad (5)$$

with $P_c^{(\beta_i)}$ the probability of correct reception of the i^{th} frame, D_{o_i} (resp. D_{loss_i}) the average GOP distortion observed with frames 0 (Intra) to i being correct (resp. observed due to the i^{th} frame being lost). Naturally, these conditional probabilities could be more precisely attuned should one dispose of more information on the dependencies between frames (e.g. reference frame numbers for each frame).

Considering again the example of a memoryless erroneous channel with bit error probability P_e , the probability of correct reception is $P_c^{(\beta_j)} = (1 - P_e)^{\beta_j \cdot n_j}$, yielding:

$$\begin{aligned} \hat{D}_{gop} &= \prod_{i=0}^N (1 - P_e)^{\beta_i n_i} \cdot D_o + \\ &\sum_{i=0}^N \left[\prod_{j=0}^{i-1} (1 - P_e)^{\beta_j n_j} \cdot (1 - (1 - P_e)^{\beta_i n_i}) \cdot D_{loss_i} \right] \end{aligned} \quad (6)$$

with $D_o = D_{o_N}$ the average GOP distortion without error.

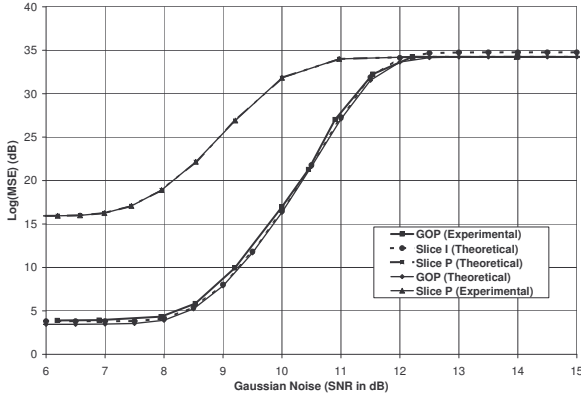


Fig. 2. First GOP ($I_1 P_{14}$) average PSNR sensitivity of 'Foreman' sequence.

Semi-analytically derived with estimating the different frames sizes and distortions resulting of correct transmission and of each frame loss, the sensitivity obtained for H.264/AVC encoded sequence 'Foreman' first GOP in QCIF format, $QP_I = 28, QP_P = 30$ is very close to the simulation results obtained for the PSNR average sensitivity (over 6000 trials) when transmitting said sequence over an AWGN channel, as shown in Fig 2. One also note that the overall distortion is mostly impacted by the most sensitive frame.

C. Data Partitioning

When the stream is data partitioned, each P frame is carried over up to three slices (NAL-A, NAL-B, NAL-C), with each slice depending on the same frame previous ones for correct decoding. To take into account slice dependency, we will

then assume that should a partition be lost, the distortion resulting from a later partition badly received is negligible. The sensitivity of a DP GOP is then deduced from Eq. (6):

$$\begin{aligned} \hat{D}_{gopDP} &= \prod_{i=0}^N \prod_{k=1}^3 (1 - P_e)^{(1 - \beta_{i,k}) \cdot n_{i,k}} \cdot D_o + \\ &\sum_{i=0}^N \sum_{k=1}^3 \left[\prod_{j=0}^N \prod_{\ell=1}^{k-1} (1 - P_e)^{\beta_{j,\ell} n_{j,\ell}} \prod_{j=0}^{i-1} (1 - P_e)^{\beta_{j,k} n_{j,k}} \right. \\ &\quad \left. \left(1 - (1 - P_e)^{(1 - \beta_{i,k}) \cdot n_{i,k}} \right) D_{loss_{i,k}} \right] \end{aligned} \quad (7)$$

with $n_{i,k}$ the length of i^{th} frame k^{th} partition, leading to a distortion $D_{loss_{i,k}}$ if said partition is lost, and $n_i = \sum_{k=1}^3 n_{i,k}$.

The different slices sensitivity estimations for first GOP of sequence 'Foreman' in QCIF format encoded in H.264/AVC DP mode are given in Fig. 3, with the Intra frame coded over a unique slice. It can be seen that the simulation results, obtained with $QP_I = 27, QP_P = 30$ over 6000 trials are very close to the analytical expression.

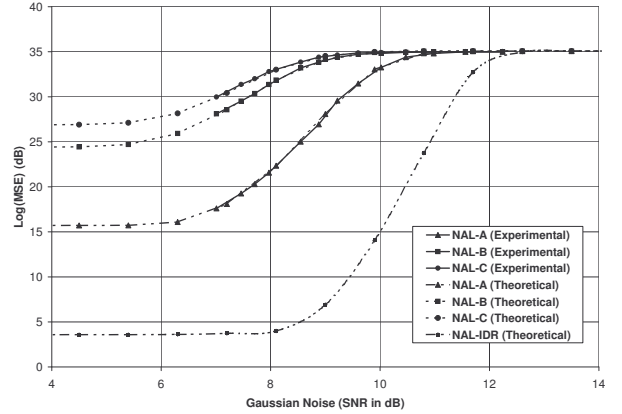


Fig. 3. 'Foreman' first GOP ($I_1 P_{14}$) average PSNR sensitivity in DP mode.

III. NUMERICAL RESULTS OVER WIRELESS CHANNEL

A. Introducing error correction by means of RCPC

A very convenient way to provide different levels of protection for different parts of a same bitstream is to adapt the protection rate by means of RCPC codes [12]. Almost as efficient as the best known convolutional codes of same protection rate, these codes offer a low complexity and allow to reach different coding rates thanks to pre-defined puncturing tables, offering an error event probability over an AWGN channel bounded by [12]:

$$P_e \leq \frac{1}{P} \sum_{d=d_{free}}^{\infty} a_d \cdot P_d \quad (8)$$

with d_{free} the code free distance, a_d the number of existing paths, $P_d = \frac{1}{2} \text{erfc} \left(\sqrt{\frac{d \cdot E_S}{N_o}} \right)$ the probability that the wrong path at distance d is selected when $SNR = E_S/N_o$.

Easily enough, the distortion of an H.264/AVC encoded sequence transmitted over a Gaussian channel and protected with an RCPC code can consequently be estimated by using this P_e value in the established distortion expressions.

B. Choosing the best protection/compression trade-off

A first application of the semi-analytical expressions established in Section II is to select the best trade-off between protection and compression for a given working point, by comparing the sensitivities resulting from the different configurations of source and channel coding for a global fixed bitrate over the channel. This is illustrated by Fig. 4 where analytical and simulated sensitivities obtained for sequence 'Foreman' in QCIF format with different compression/protection rates for a channel bitrate of 64 kbps are plotted. One sees that our models emulate quite well the simulations, and that the configuration providing the best resulting PSNR for a given working point can easily be determined. For instance, for $SNR = 3dB$, the best configuration among the proposed ones is to encode the video sequence at 21.3 kbps, then protect it with code rate $1/3$, providing more than 5 dB gains in terms of PSNR when compared with the other possible configurations.

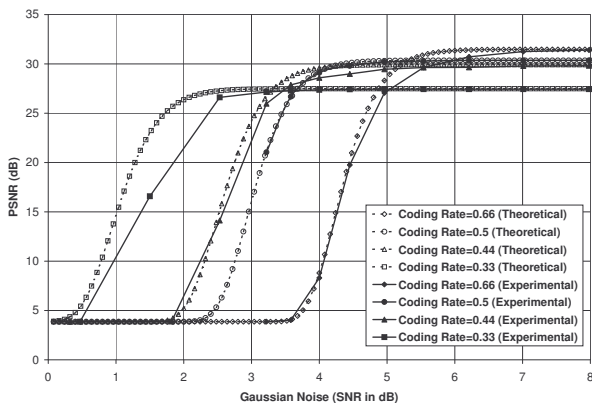


Fig. 4. 'Foreman' sequence first GOP (I_1P_{14}) average PSNR sensitivity for different configurations with channel bitrate 64 kbps.

C. Unequal Error Protection in Data partitioning mode

Another application for the established semi-analytical expressions is the determination of the different protection rates to be applied in an unequal error protection (UEP) context, in particular when the H.264/AVC codec is working in Data partitioning mode. Indeed, the different partitions, from the NAL-IDR to the NAL-C one, have different sensitivities, as illustrated in Fig. 3. Using the corresponding overall distortion expression given in Equation (7), it is possible to choose the best parameters of puncturing rate of RCPC for each partition by comparing the resulting expected distortion with the different configurations of coding parameters. As an example, Figure 5 gives the results obtained for both 'Foreman' (64 kbps, $R = 1/2$) and 'Stefan' (185 kbps, $R = 2/3$) sequences, with same protection rate R in equal error protection (EEP) and UEP modes. For both sequences, including 'Stefan' whose motion level is higher than 'Foreman', gains of 5 to 10 dB in terms of PSNR are obtained when compared to EEP mode.

IV. CONCLUSIONS

Semi-analytical expressions for the distortion due to video transmission over a wireless channel for Intra, Predicted frames, GOPs and data partitioned GOPs have been presented in this paper. The proposed models take into account the

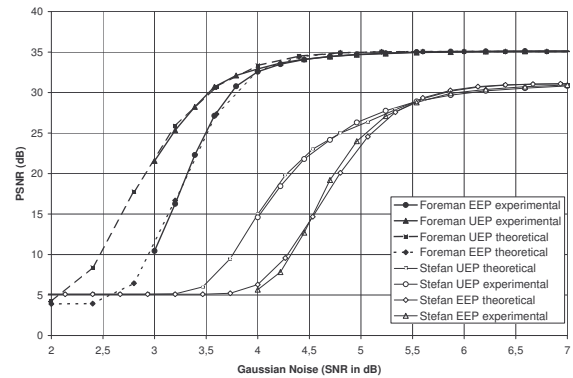


Fig. 5. 'Foreman' and 'Stefan' first GOP (I_1P_{14}) average PSNR in EEP/UEP. effects of error propagation within the frame or/and the GOP, as well as the impact of the lossy H.264/AVC compression. It was also shown that the model remained valid when used together with error protection tools, whether equal or unequal ones to give expression of the overall source and channel distortion D_{S+C} also when in presence of channel coding.

Simulation results show the accuracy of the models, and illustrate their interest to determine the best trade-off of channel and source coding rates for a given operating point, or to select the partitions protection rates when considering Unequal Error Protection. Over an AWGN channel, gains of 5 to 10 dB of PSNR are obtained when compared to not optimised solutions. Future works include the derivation of the model in the context of hierarchical frames.

ACKNOWLEDGMENT

This work was partially supported by the European Community with project IST-FP6-001812 PHOENIX.

REFERENCES

- [1] C.E. Shannon, "A Mathematical Theory of Communication," in *The Bell System Technical Journal*, pp. 379-423, pp. 623-656, July, Oct. 1948.
- [2] N. Kamaci, Y. Altunbasak, and R.M. Mersereau, "Frame Bit allocation for the H.264/AVC video coder via Cauchy-Density-Based rate and distortion models" *IEEE Trans. Circ. Syst. for Video Tech.*, Aug. 2005.
- [3] S. Wenger, "H.264/AVC over IP," *IEEE Trans. Circ. Syst. for Video Tech.*, vol. 13, n. 7, pp.645-656, July 2003.
- [4] J. Shin, J. Kim, and C-C.J. Kuo, "Quality-of-service mapping mechanism for packet video in differentiated services network," in *IEEE Trans. on Multimedia*, vol. 3 n. 2, pp.219-231, June 2001.
- [5] J. Chakareski and P. Frossard "Low-complexity Adaptive streaming via optimized A priori Media Pruning", in *Proc. Int. Workshop on Multimedia Processing (MMSP'05)*, Shanghai, China, Oct-Nov 2005.
- [6] Z. He, J. Cai and C.W. Chen "Joint source channel rate-distorsion analysis for adaptive mode selection and rate control in wireless video coding", in *IEEE Trans. Circ. Syst. for Video Tech.*, June 2002.
- [7] M. Bystrom and T. Stockhammer, "Dependent source and channel rate allocation for video transmission", in *IEEE Trans. on Wireless Comm.*, vol. 3, n. 1, pp. 258-268, Jan. 2004.
- [8] M.G. Martini and M. Chiani, "Rate-Distortion models for Unequal Error Protection for wireless video transmission", in *Proc. IEEE Vehicular Technology Conference (VTC'04)*, pp. 1049-1053, 2004.
- [9] C. Lamy-Bergot, N. Chautru and C. Bergeron, "Unequal Error Protection for H.263+ bitstreams over a wireless IP network", in *Proc. of the IEEE ICASSP'06 conference*, pp. V-377/V-380, Toulouse, France, May 2006.
- [10] Joint verification model for H.264 (JM10.1), <http://iphone.hhi.de/suehring/tml>, Nov. 2005.
- [11] C. Bergeron and C. Lamy-Bergot, "Compliant selective encryption for H.264/AVC video streams", *Proc. Int. Workshop on Multimedia Processing (MMSP'05)*, pp. 477-480, Shanghai, China, Oct-Nov 2005.
- [12] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their application," in *IEEE Trans. on Comm.*, vol. 36, n. 4, pp. 339-400, April 1988.