Error resilient H.264/AVC video over satellite for low packet loss rates

Shankar Manuel Aghito, Søren Forchhammer and Jakob D. Andersen
COM, DTU Department of Communications, Optics and Materials
Ørsteds Plads 343, 2800 Kgs. Lyngby, Denmark
sma@com.dtu.dk sf@com.dtu.dk jda@com.dtu.dk

Abstract

The performance of video over satellite is simulated. The error resilience tools of intra macroblock refresh and slicing are optimized for live broadcast video over satellite. The improved performance using feedback, using a cross-layer approach, over the satellite link is also simulated. The new Inmarsat BGAN system at 256 kbit/s is used as test case. This systems operates at low loss rates guaranteeing a packet loss rate of not more than $10^{-3}$. For high-end applications as ‘reporter-in-the-field’ live broadcast, it is crucial to obtain high quality without increasing delay.

1 Introduction

Using satellite communication, video may be shot at the most remote areas and used e.g. for live broadcast. Considering such a ‘reporter-in-the-field’ application, the satellite communication will constitute the bottle-neck. Recently, Inmarsat has introduced the Broadband Global Area Network (BGAN) [2], which offers 256 kbit/s satellite communication using portable terminals. For the high-end live broadcast application, the video may be visualized on high-definition flat panel displays for nation-wide scrutiny of the video signal. We have simulated the coding parts of the system, i.e. the basic Turbo coding used for forward error-correction (FEC) which uses FEC packets of 5, 10 and 20 ms at different signal-to-noise ratios. The resulting packet-loss rates are used to analyse the packet-loss effect on H.264 [4] video using the H.264 reference software. In this work we focus on the video coding side. Feedback over the satellite link based on geo-stationary satellites is possible, but with an inevitable delay. Using this delayed feed-back is also simulated. The error resilience tools of the H.264 reference software considered are intra macro block refresh and slicing, which are suitable for low delay transmission using portable devices, since they do not impose requirements on the complexity of the encoder and do not introduce additional delay. A recent review of error resilience tools for H.264 is given in [1].

This work was supported in part by the Danish Agency for Science Technology and Innovation.

2 Simulation Set-up

In H.264 encoded video, pictures are partitioned in one or more slices. Each coded slice is embedded into one Network Adaption Layer Units (NALU). We assume that each NALU is inserted into one RTP packet. The conventional RTP/UDP/IP protocol stack is utilized together with robust header compression (ROHC), which typically compresses header data into 3 bytes [6]. The whole set-up is simulated simply using the H.264 reference software with Annex B output format (4 bytes start code in each slice). The IP stream is fed to the BGAN, which divides the data into FEC packets (not synchronized with the NALUs). The FEC packets are divided into packet data units (PDU). Damaged FEC packets are detected by a CRC-check on the PDUs. Initial simulations indicated that discarding a whole FEC packet will be the dominating case even if this hold multiple PDUs. (The probability of some PDUs being correct when the FEC packed is damaged is small.) Thus the model chosen is to discard all NALUs which are hit by a discarded FEC packet. We consider an average (FEC) packet loss rate (PLR) of $10^{-3}$, whereas most of the work available in the literature has focussed on considerably higher PLR according to the common test conditions defined in [5],[6]. For the satellite channel at this low level of loss it is reasonable to assume the loss of packets to be independent.

Even for a given coded video stream the effect of loosing different NALUs may lead to large differences in the resulting decoded video after error concealment. Considering the loss of packets to be independent, we further assume that the individual (FEC) packets lost do not interact. (For one lost FEC packet leading to one or more lost NALUs we assume recovery prior to the next loss and/or the effect of these to be additive as in [3].) Hence the effect of single FEC packet losses on the decoded video is analysed. To capture the variations, the FEC packets of one coded sequence are lost in turn and statistics gathered on the (loss) performance.

3 Simulation Results

The sequences utilized are carphone, foreman, and mother and daughter (MD) (30Hz, CIF resolution and 382, 400, and 400 frames, respectively). The H.264 reference
A first experiment was performed in order to evaluate the impact of the error resilience tools on the average PSNR, in the error-free case. All combinations of IMBR (RandomIntraMBRefresh (IMBR) and SliceArgument (with SliceMode = 1, so that Slices = 396/SliceArgument), respectively. In the decoder, error concealment of entire pictures is done in the ‘motion copy’ mode.

A second experiment was aimed at analyzing the impact of slices per picture on the number of IMBR and slices per picture is linear (carphone).

We note the leaky nature of the loss extending beyond the point where all macro-blocks have been intra updated. For the static sequence MD the corresponding decrease is about 0.55 dB. The larger penalty in the latter case is expected, since static sequences are encoded more efficiently if inter prediction is fully utilized.

A second experiment was aimed at analyzing the impact of the error resilience tools on the average PSNR around 0.08 dB. The effect of IMBR depends on the amount of motion: for the fast motion sequences carphone and foreman, increasing the value of IMBR by 5 the PSNR is reduced about 0.3 dB, while for the static sequence MD the corresponding decrease is about 0.55 dB. The larger penalty in the latter case is expected, since static sequences are encoded more efficiently if inter prediction is fully utilized.

A second experiment was aimed at analyzing the impact of the loss of single packets. The experiment was carried out by removing FEC packets of 5, 10 and 20 ms (BGAN [2]) from the bitstream. One packet at the time is assumed to be corrupted, all the NALUs hit by the erroneous packet are removed, and the effect on the PSNR (i.e. the loss compared to the error-free case) is captured in an interval of 3 s from the occurrence of the error (i.e. 90 frames). The results for packets of 10 ms and settings IMBR=10 and Slices=9 for the sequence carphone are displayed in Fig. 2. As a first approximation, the average PSNR loss may be modelled as an exponentially decreasing function. The initial loss depends on the number of slices and the size of the packet, while the slope of the exponential depends on IMBR (both initial loss and slope depend of course also on the type of sequence). We note the leaky nature of the loss extending beyond the point where all macro-blocks have been intra updated. For the packet sizes of 5, 10, and 20 ms, given PLR = 10^{-3}, errors occur in average every 150, 300, and 600 frames. Results are reported in Table 1 comparing settings that provide roughly the same error-free PSNR, for each sequence (in two cases packet lengths are assumed equally likely, since in the BGAN these cannot be controlled by the user and the authors do not have better statistics). These were chosen in order to: maintain the average PSNR, with average losses compared to the non resilient setting (IMBR=0, Slices=1) between 1 and 2 dB; keep the residual average PSNR loss after 90 frames from the error well below 1 dB, in order to validate the assumption that the loss of one packet is recovered before the occurrence of the next error.

An initial evaluation, aimed at relating the PSNR loss and the perceived duration of error propagation to the overall perceived quality is shown in Fig. 3. The displayed cases refer to observable effects of the loss of data, i.e. errors generating peak PSNR losses smaller than 2 dB were not
A last experiment was carried out for analyzing the improvement that can be obtained by using a feedback channel in order to signal to the encoder the loss of one packet. The feedback delay was assumed to be 500 ms. A simple solution was simulated, by allowing the encoder to switch to a new setting, named recovery mode, when the error is signaled. The recovery mode utilizes a higher value of IMBR in order to speed-up the recovery, and only 1 slice per picture such that the average PSNR (error-free) is maintained. An example is shown in Fig. 4, where the simulated loss profile is obtained by combining the profiles of the two settings. Results for the three sequences are reported in Table 1. By using the recovery mode the average PSNR is slightly improved, and a significant reduction of the PSNR loss after 500 ms is observed. This reduction should provide an important benefit in terms of visual quality evaluation (see also Fig. 3).

The setting IMBR/Slices = 10/9 is a robust solution over the different packet sizes, with good balance between error localization and recovery, providing the best average PSNR results for foreman and carphone (Table 1). For the less critical sequence MD, slightly better average PSNR are obtained with setting 15/1, since in static sequences error concealment works well even when a whole picture is lost.

Simulation of the video coding when loosing a FEC packet of 5, 10 or 20 ms has shed light on the influence on the video performance. Based on the simulations of the Turbo FEC system it is possible to include the effect on the FEC efficiency. By introducing various puncturing schemes in the Turbo coding the FEC rate may be changed. Consider the performance for the 10 ms packets and fix the signal-to-noise (SNR) at the level corresponding to a packet loss-rate of $10^{-3}$. At this SNR level change the block size and rate of the Turbo coding, increasing the rate to 263 kbit/s for 20 ms packets and decreasing it to 248 kbit/s for the 5 ms packets. This leads to changes in the PSNR of the coded video. For foreman 20 ms FEC leads to a 0.15 dB increase and 5 ms FEC to a 0.15 dB decrease compared with 10 ms FEC. The PSNR values of Table 1 could be adjusted accordingly.

4 Conclusions

An analysis of the quality of H.264/AVC video at low packet loss rates over BGAN was presented. The effect of intra-macroblock refresh and slicing was analysed, in terms of average and instantaneous PSNR values, for different FEC packet sizes. A robust setting for all packet sizes was found. The benefit of a (500 ms) delayed feedback channel was demonstrated.

Acknowledgements

We would like to thank Thrane & Thrane A/S for the fruitful discussions on the topic.

References

Table 1. Objective quality evaluation on carphone (top), foreman (center) and mother and daughter (bottom), for different IMBR/Slices settings. The average error free PSNR values without using error resiliency, i.e. for the setting 0/1, are 34.05, 33.52 and 39.84, respectively. Mean PSNR and instantaneous PSNR loss (at \{0, 0.5, 1, 2, 3\} seconds after error occurrence) are reported. Values in ( ) indicate the quality obtained using recovery mode. For foreman and MD results are reported assuming equally likely packet lengths (the relative performance between different packet lengths is similar as that provided for carphone).