VIDEO CODING AND TRANSPORT LAYER TECHNIQUES FOR H.264/AVC-BASED TRANSMISSION OVER PACKET-LOSSY NETWORKS

Thomas Stockhammer
Institute for Communications Engineering
Munich University of Technology
80290 Munich, Germany

Thomas Wiegand
Image Processing Department
Heinrich-Hertz- Institute,
10587 Berlin, Germany

Tobias Oelbaum, Florian Obermeier
Institute for Data Processing
Munich University of Technology
80290 Munich, Germany

ABSTRACT

Standard compliant enhancements of the emerging H.264/AVC coding algorithm for transmission over lossy IP-based networks are described with the error resilience features being presented in greater detail. Standard compliant encoder and decoder enhancements as well as the exploitation of transport layer mechanisms to improve the quality in packet lossy environments are presented. Different schemes are compared and appropriate experimental results based on common test conditions are discussed. Finally, the selection of suitable features is discussed.

1 INTRODUCTION

H.264/AVC [1] is an attractive candidate for many applications including fixed and wireless video transmission over the Internet Protocol (IP) due to its significantly increased compression efficiency compared to state-of-the art video coding standards such as H.263 or MPEG-4. However, to allow transmission in different environments not only coding efficiency is relevant, but also seamless and easy integration of the coded video into all current and possible future protocol and multiplex architectures and enhanced error resilience features are of major importance. Among others, especially IP-based standard compliant video transmission has attained significant interest recently. Typical applications include conversational services, such as video telephony, and videoconferencing, streaming services or multimedia messaging services. In addition to traditional fixed Internet video services, also the video transmission over emerging and future mobile systems will be mainly packet-based.

For real-time video such as conversational or streaming services usually IP on the network layer, user datagram protocol (UDP) on the transport layer, and real-time transport protocol (RTP) and accompanying RTP payload specifications are employed. However, UDP offers only a simple, unreliable datagram transport service: packets may get lost, duplicated, or re-ordered on their way from the source to the destination due to network congestion, buffer overflows in intermediate routers or frame losses on mobile links.

Especially, for conversational services without retransmission possibilities, the highly complex temporal and spatial prediction mechanisms included in modern video codecs like H.264/AVC result in catastrophic error propagation in case of packet losses. Then, the use of error resilience techniques in the source codec becomes important. Many schemes addressing this issue have been previously presented, investigated and assessed, e.g., see [2]-[7] and references therein. The prime goal of this work is the adaptation of well-known and successfully applied techniques to H.264/AVC. We formulate the basic problems, present and evaluate H.264/AVC error resilience features and transport layer mechanisms.

2 H.264/AVC IN IP-BASED ENVIRONMENT

2.1 Problem Formulation

The investigated video transmission system is shown in Figure 1. H.264/AVC video encoding is based on a sequential encoding of frames denoted with the index $n=1,\ldots,N$ with N the total number of frames to be encoded. In most existing video coding standards including H.264/AVC, within each frame video encoding is typically based on sequential encoding of macroblocks denoted by index $m=1,\ldots, M$ where M specifies total number of macroblocks in one frame and depends on the spatial resolution of the video sequence. The encoding process can form slices by grouping a certain number of macroblocks.

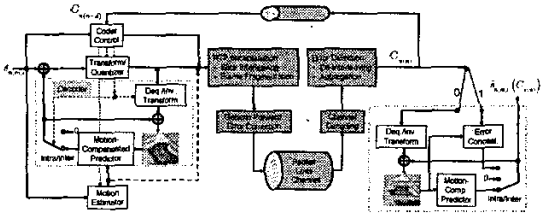

Figure 1: H.264 coding in IP environment with RTP encapsulation, forward error correction and delayed feedback information.

The generated slices are mapped to Network Adaptation Layer (NAL) units. The RTP payload specification [8] specifies simple encapsulation of NAL units. In addition, several NAL units can be combined into one aggregation packet or one NAL unit can be fragmented into several transport packets. Applications will be discussed in Section 3.

For notational convenience, let us define the number of transmission packets to transmit all frames up to n as $\pi(n)$. With that, we can define the packet loss or channel behavior c as a binary sequence $\{0,1\}^{\pi(n)}$ indicating whether a slice is lost (indicated by 1) or correctly received (indicated by 0). Obviously, if a NAL unit and the encapsulated slice is lost all macroblocks contained by this slice are lost. It can be assumed that the decoder is aware of any lost packet due to the error detection in the underlying system. The channel loss sequence is obviously random and, therefore, we denote it as $C_{\pi(n)}$ where the statistics are in general unknown to the encoder. According to Figure 1, in addition to the forward link it is possible that a low bit-rate reliable back-channel from the decoder to the encoder is available which allows reporting a d-frame delayed version of the observed channel behavior at the decoder $C_{\pi(n-d)}$ to the encoder. In RTP/IP environments this is usually based on RTCP messages.

The decoder processes the received sequence of packets. Whereas correctly received packets are decoded as usual, for the lost packet an error concealment algorithm has to be invoked. The reconstructed sample $\hat{s}_{n,m,i}$ at position i in macroblock m and frame n depends on the channel behavior and on the decoder

0-7803-7750-8/03/$17.00 ©2003 IEEE

error concealment. In Inter mode, i.e., when motion-compensated prediction (MCP) is utilized, the loss of information in one frame has a considerable impact on the quality of the following frames, if the concealed image content differs from the decoded content in case of no errors and this content is referenced for MCP. Therefore, due to the motion compensation process and the resulting error propagation, the reconstructed image depends not only on the lost packets for the current frame but in general on the entire channel loss sequence $C_{\pi(n)}$. We denote this dependency as $\hat{s}_{n,m,i}\left(C_{\pi(n)}\right)$.

From this system perspective an error-resilient video coding standard suitable for conversational IP-based services has to provide features to combat various problems, always focusing on prime goal of high compression efficiency. The tools required in an error-prone environment can be divided into the following major categories:

1. Reduction of errors that result in packet loss using the selection of slice sizes and channel coding techniques such as forward error protection (FEC)
2. Concealment of errors in the decoded picture in case the channel coding techniques failed.
3. Mitigation of spatio-temporal error propagation that is caused when the concealed samples and the samples that would have been decoded are different and these samples are referenced for motion compensation.

Before discussing H.264/AVC standard features such as test model extensions for encoder and decoder as well as transport issues which related to the discussed problems, we will briefly present the applied test conditions.

2.2 Common Test Conditions

The H.264/AVC standardization process acknowledged the importance of IP-based transmission by adopting a set of common test conditions for IP based transmission [9]. These conditions allow selecting appropriate coding features, testing and evaluating error resilience tools, and producing meaningful anchor results. Anchor video sequences, appropriate bit-rates and evaluation criteria are specified. We will present results for a representative selection of the common Internet test conditions[1]. The applied test case combinations include the QCIF sequences Foreman and Hall Monitor as well as the CIF sequence Paris, all with an original frame rate of 30 frames per second (fps). The first 300 frames of the original sequence are encoded at a frame rate of 7.5 fps for Foreman and 15 fps for Hall Monitor and Paris applying only temporally backward referencing motion compensation.

As some error resilience features (e.g. slice coding) were not updated in later versions of the H.264 test model software, we decided to use JM1.7 for these experiments. This version does not include a rate control, therefore we chose to present the results when encoding the sequence with a fixed quantization parameter. For all test results the sequences were encoded with the quantization parameter *q=12, 16, 20, 24, 28* (according to WD-2) and measured the resulting total bit rate including a 40 byte IP/UDP/RTP header for each transmitted packet. All tests are carried out with five reference frames. As performance measure we chose the commonly applied averaged PSNR of the luminance component (Y-PSNR) where the average is taken over all encoded frames. To obtain sufficient statistics we transmitted at least 3000 frames for all experiments, using a simple packet loss simulator. The applied Internet error patterns captured from real-world measurements result in a packet loss rates of approximately 3%, 10%, and 20%, respectively.

[1] For a complete set of results, see http://www.ei.tum.de/~stockhammer.

3 ERROR RESILIENCE IN H.264 – FEATURES AND EXPERIMENTAL RESULTS

3.1 Macroblock Intra-Updates

Although we will present some techniques which allow reducing the packet loss rate or at least the visual effects of these losses, transmission errors and resulting reference frame mismatches between encoder and decoder are usually not avoidable. Then, the effects of spatio-temporal error propagation are in general severe. A quick recovery can be achieved when image regions are encoded in Intra mode, i.e., without reference to a previously coded frame. Completely Intra coded frames are usually not inserted in real-time and conversational video applications as the instantaneous bit-rate and the resulting delay is increased significantly. Instead, H.264/AVC allows encoding of single macroblocks for regions that cannot be predicted efficiently as it is also known from other standards. Another feature in H.264/AVC is the possibility to select the reference frame from the multi-frame buffer. Both features have mainly been introduced for improved coding efficiency, but they can efficiently be used to limit the error propagation. Conservative approaches transmit a number of Intra coded macroblocks anticipating transmission errors. In this situation, the selection of Intra coded macroblocks can be done either randomly or preferably in a certain update pattern. For details and early work on this subject we refer to, e.g. [10].

The selection of appropriate coding options in many source-coding standards in order to improve coding efficiency is often based on rate-distortion optimization (CE-RDO) algorithms [11]. The extension of this approach to error-prone transmission has been proven to perform significantly better than heuristic methods [12]-[14], i.e. channel-adaptive rate distortion optimization (CA-RDO). A key issue is the estimate of the expected sample distortion in packet loss environments which has been addressed in several papers. For the results in this work we have chosen the approach with multiple independent channel-decoder pairs in the encoder according to [15]. Figure 2 shows the RD performance for CA-RDO schemes compared to CE-RDO with pseudo-random updates with different update ratios for the 10% error pattern. It can be observed that the channel-adaptive mode selection outperforms the best pseudo random intra-update strategies. The gain of 1-2.5 dB compared to the best regular intra updates is obvious. For the same quality the bit-rate decreases for adaptive intra-updates by about 30%.

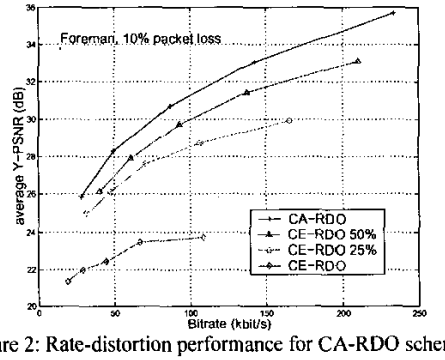

Figure 2: Rate-distortion performance for CA-RDO schemes compared to CE-RDO with pseudo-random updates with different update ratios for the 10% error pattern.

3.2 Slices and Error Concealment

Packet loss probability and the visual degradation from packet losses can be reduced by introducing slice-structured coding,

which provides spatially-distinct resynchronization points within the video data for a single frame. On the one hand, short packets reduce the amount of lost information and, hence, the error is limited and error concealment methods can be applied successfully. In the H.264/AVC test model the simple previous frame copy (PFC) error concealment has been replaced by advanced error concealment (AEC). On the other hand, the loss of spatial prediction within one frame and the increased overhead associated with decreasing slices adversely affect performance.

The RTP payload specification of H.264/AVC includes the concept of aggregation packets, which means that several NAL units can be transported within one IP packet. This allows the concept of *Slice Interleaving* [17]. For our simulations we applied that all slices containing odd macroblock rows are transmitted within the first IP packet and all slices containing even macroblock rows are transmitted within the second IP packet. This concept does not reduce the coding overhead due to the limited spatial prediction, but the costly IP overhead of 40 bytes per packet can be avoided.

A more advanced and generalized concept is given by *flexible macroblock ordering* (FMO) [18] providing the possibility to transmit macroblocks in non-scan order. This flexibility allows the definition of different patterns - including slice interleaving - without interrupting the inter macroblock prediction for motion vector prediction and entropy coding. FMO is especially powerful with appropriate error concealment.

A third error resilience concept in H.264/AVC is *data partitioning*, which can also reduce visual artifacts resulting from packet losses, especially if prioritization or unequal error protection is provided by the network. In this paper we will not further investigate FMO and Data Partitioning, since this is subject of current and future work.

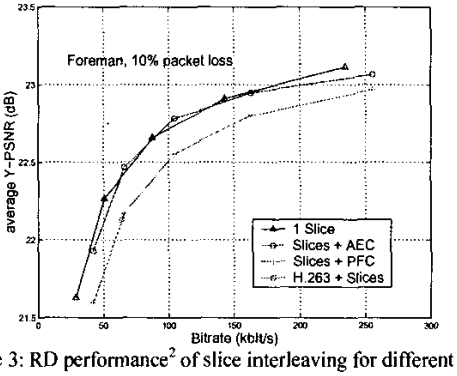

Figure 3: RD performance[2] of slice interleaving for different concealments schemes compared to single slice scheme and to H.263+ with slice interleaving, (Foreman, 10%).

Figure 3 shows the RD performance[2] of slice interleaving with two IP packets per video frame interleaving for different concealments schemes compared to single slice scheme and to H.263+ with slice interleaving, (Foreman, 10%). For both cases channel-adaptive RD optimization has been applied. Different error concealment strategies are assessed for slice interleaving and the results show that a significant gain for AEC compared to the PFC is visible. However, in general the introduction of slices is not beneficial compared to single slice approach as the interruption of the spatial prediction reduces the coding performance significantly. In addition, with more packets per picture the probability that a certain video picture is affected by an error increases if we assume that the loss rate is independent of the packet length. This might be different if packet losses are caused by bit errors such as in mobile environments. In comparison to optimized H.263+ [19] H.264/AVC performs better in case of advanced error concealment and identical with simple previous frame copy.

[2] Note that in this case a different PSNR measure has been used. The decoded sequence is compared to each and every frame of the original sequence. The results for H.263+ have been obtained in this way.

3.3 Forward Error Correction

In general any kind of FEC in combination with interleaving for packet lossy channels can be applied. A simple solution is provided by RFC2733 [20], more advanced schemes have been introduced, e.g. in [21]. In the following we will apply a simple FEC scheme for each frame. Therefore, we assume that each frame is fragmented into k_f packets of equal size. An appropriate syntax and semantics to JVT has been proposed and the fragmentation scheme has been added to the RTP payload specification [8] just recently. Based on the Reed-Solomon code properties and RFC2733 a simple FEC scheme can be constructed such that if we receive just any k_f packet out of n_f transmitted packets, we can reconstruct the entire k_f information packets and, therefore, the entire frame.

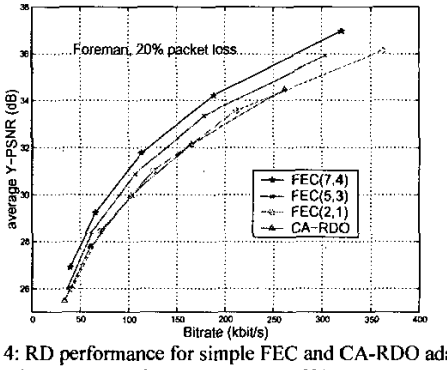

Figure 4: RD performance for simple FEC and CA-RDO adapted to 3% loss rate using Foreman and 20% error pattern

Figure 4 shows the RD performance for simple FEC and CA-RDO adapted to 3% loss rate using Foreman and 20% error pattern. We assume, that in case that there are less than k packets received for one frame, the entire frame is lost. In addition, we apply the same error protection, specified by (n, k), for all frames of the sequence. We target for all FEC schemes for the same residual error rate of approximately 3%. Obviously, there exists an optimum number of information packets per frame k_f, as for too small n, the code is very weak and for too high n, the packetization overhead limits the performance. It can be seen that especially for high loss rates (20 %) significant gains with FEC can be achieved when compared to the optimized intra update scheme without FEC. For the 10 % error rate case the gains are less significant, but for higher data rates they are still remarkable. For lower error rates we have found that FEC does not provide noticeable gains compared to optimized intra updates.

3.4 Multiple Reference Frames and Feedback

Multiple reference frames can be used to limit the error propagation, for example as has been shown for H.263++ in [13] or in *video redundancy coding* schemes [22]. Moreover, multiple reference frames can be combined with a feedback channel [13]. So far we have assumed that there is no feedback information from the decoder except for a possible report of the average packet loss rate to adapt the intra updates in the macroblock mode selection. However, the knowledge of a d-frame delayed version of the observed channel characteristic at the encoder might be useful even if the erroneous frame has already been decoded. This characteristic can be conveyed from the decoder to the encoder by

acknowledging correctly received transport packets (ACK), sending a not-acknowledge message (NACK) for missing packets or both types of messages. In general it can be assumed that the reverse channel is error-free and the overhead is negligible. In previous standards such as H.263 or MPEG-4 this feedback has been exploited for Error Tracking [23] or simple NEWPRED techniques [24]. Novel uses of this technique have been described in [13] and also carry over to H.264/AVC as shown below.

The flexibility provided in H.264/AVC coding to select macroblock mode and reference frames on macroblock and sub-macroblock basis allows incorporating NEWPRED techniques in a straight-forward manner [13]. Results with two different approaches are presented. *ACK mode* allows to reference only acknowledged reference area, otherwise intra coding is applied. If the encoder is aware of the error concealment in the decoder, it can apply the same error concealment for the reference frame for the non-acknowledge area (NACK). Both schemes are combined with the CE-RDO. They allow eliminating the drift of reference frames and error propagation.

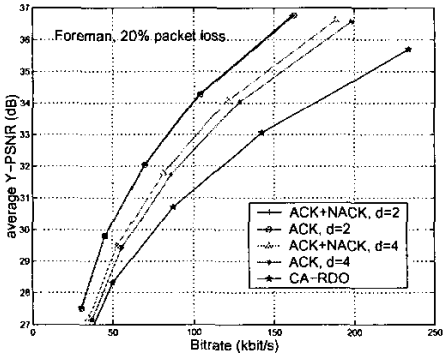

Figure 5: RD performance for feedback methods with different feedback delays d compared CA-RDO for Foreman, 10%.

Figure 5 shows the RD performance for feedback methods with different feedback delays d compared to CA-RDO for Foreman and a 20% packet loss rate. Significant gains compared to CA-RDO can be observed for low feedback delays. For increasing feedback delays the performance obviously decreases as the reference frames available are further in the past and, therefore, in general less correlated to the actual frame. Unless the number of reference frames is just one more than the feedback delay (see $d=4$), the ACK only mode is similar to the mode which includes NACK and error concealed area at the decoder.

4 CONCLUSIONS

In this paper we discussed standard compliant enhancements of the emerging H.264/AVC coding algorithms for transmission over lossy IP-based networks. The error resilience features within the H.264/AVC coding algorithm have been presented and assessed in combination with transport layer mechanisms in greater detail. It is worth to mention that for the results presented in this paper the objective results based on the average PSNR match the observed subjective quality quite well. The presented results and additional experiments based on the test conditions[1] allow drawing the following conclusions.

Unless slice structured coding is necessary to limit the transport packet layer size, it may be preferable to transport one frame in one transport layer packet for sequences like the tested one. The reduction in coding efficiency due to limited spatial prediction and increased packet overhead can not be compensated by improved error concealment at the decoder. However, if slice structured coding is applied, advanced error concealment does provide significant gains.

An efficient method to limit error propagation seems to be multiple reference frames combined with low-delay feedback information. If no feedback channel is available CA-RDO show very good performance. If the feedback delay is higher, then a combination of CA-RDO and feedback approaches might be beneficial [13], [14]. Only for packet error rates of about 10% or higher the application of FEC schemes combined with CA-RDO seems to be interesting. The combination of FEC with feedback approaches is also currently being investigated.

5 REFERENCES

[1] T. Wiegand (ed.), "Committee Draft No 1, Rev. 0 (CD-1)," JVT-C167, May 2002.

[2] P. Bahl and B. Girod (Eds.), "Wireless Video", vol. 36, IEEE Communcations Magazine, June 1998.

[3] J. C. Brailean, T. Sikora, and T. Miki (Eds.), "Special Issue on Error Resilience", vol. 14, Signal Processing: Image Communication, May 1999.

[4] H. Gharavi and L. Hanzo (Eds.), "Special Issue on Video Transmission for Mobile Applications", Proc. IEEE, Oct. 1999.

[5] A. Reibman and M. T. Sun (Eds.), "Wireless Video", Marcel Dekker, 2000.

[6] C. W. Chen, P. Cosman, N. Kingsbury, J. Liang, and J. W. Modestino (Eds.), "Special Issue on Error-Resilient Image and Video Transmission", IEEE JSAC, vol. 18, no. 6, June 2000.

[7] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, „Error Resilient Video Coding Techniques", IEEE Signal Proc. Mag., pp. 61-82, July 2000.

[8] S. Wenger, T. Stockhammer, and M. Hannuksela, "RTP payload Format for JVT Video", draft-ietf-avt-rtp-h264-00.txt, Internet Draft, Sept 2002.

[9] S. Wenger, "Common Conditions for wireline, low delay IP/UDP/RTP packet loss resilient testing", ITU-T SG16 Doc. VCEG-N79r1, Sept. 2001.

[10] P. Haskell and D. Messerschmitt, "Resynchronization of Motion-Compensated Video Affected by ATM Cell Loss," Proc. ICASSP, vol. 3, pp. 545–548, 1992.

[11] G.J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression," IEEE Signal Proc. Mag., vol. 15, no. 6, pp. 74-90, Nov. 1998.

[12] G. Cote, S. Shirani, F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," IEEE JSAC, vol. 18, no. 6, pp. 952 -965, June 2000.

[13] T. Wiegand, N. Färber, K. Stuhlmüller, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction", IEEE JSAC, vol. 18, no. 6, pp. 1050-1062, June 2000.

[14] R. Zhang, S. L. Regunathan, and K. Rose, "Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience," in IEEE JSAC, vol. 18, no. 6, pp. 966-976, June 2000.

[15] T. Stockhammer, D. Kontopodis, and T. Wiegand, "Rate-Distortion Optimization for JVT/H.26L Coding in Packet Loss Environment", Proc. PVW, Pittburgh, PY, April 2002.

[16] V. Varsa, M. Hannuksela, and Y. Wang, "Non-normative error concealment algorithms," ITU-T VCEG-N62, Sept. 2001.

[17] V. Varsa and M. Karczewicz, "Slice interleaving in compressed video packetization", Proc. PVW, Forte Village, Italy, May 2000.

[18] S. Wenger and M. Horowitz, "Flexible MB Ordering – A New Error Resilience Tool for IP-Based Video", Proc. IWDC 2002, Capri, Italy, Sept. 2002.

[19] S. Wenger and G. Cotê, "Using RFC 2429 and H.263+ at low to medium bitrates for low latency applications," Proc. PVW, New York, NY, April 1999.

[20] J. Rosenberg and H. Schulzrinne, "An RTP Payload Format for Generic Forward Error Correction," RFC 2733, December 1999.

[21] G. Carle, and E.W. Biersack, "Survey of Error Recovery Techniques for IP-based Audio-Visual Multicast Applications," IEEE Network Magazine, Vol. 11, No. 6, p. 2-14, Nov. 1997.

[22] S. Wenger, G. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," IEEE CSVT, pp. 867-877, November 1998.

[23] E. Steinbach, N. Färber, and B. Girod, "Standard Compatible Extension of H.263 for Robust Video Transmission in Mobile Environments," IEEE CSVT, vol. 7, no. 6, pp. 872-881, Dec. 1997.

[24] Y. Wang and Q. Zhu, "Error control and concealment for video communication: a review," Proc. IEEE, vol. 86, pp. 974-997, 1998.