

H.264/AVC for Wireless Applications

Thomas STOCKHAMMER⁽¹⁾, and Thomas WIEGAND⁽⁸⁾

(1) Institute for Communications Engineering (LNT), Munich University of Technology (TUM), D-80290 Munich, Germany, E-mail: stockhammer@ei.tum.de

(2) Image Processing Department, Fraunhofer Institute for Telecommunications Heinrich-Hertz-Institut (HHI) D-10587 Berlin, Germany, E-Mail: wiegand@hhi.de

Abstract

Video transmission in wireless environments is a challenging task calling for high compression efficiency as well as a network friendly design. These have been major goals of the H.264/AVC standardization effort addressing “conversational” (i.e., video telephony) and “non-conversational” (i.e., storage, broadcast, or streaming) applications. The video compression performance of the H.264/AVC Video Coding Layer typically provides a significant improvement. The network-friendly design goal of H.264/AVC is addressed via the Network Abstraction Layer that has been developed to transport the coded video data over any existing and future networks including wireless systems. The main objective of this paper is to provide an overview over the tools which are likely to be used in wireless environments and discusses the most challenging application, wireless conversational services in greater detail. Appropriate justifications for the application of different tools based on experimental results are presented.

1 Introduction

Since 1997, the ITU-T's Video Coding Experts Group (VCEG) has been working on a new video coding standard with the internal denomination H.26L. In late 2001, the Moving Picture Expert Group (MPEG) and VCEG decided to work together as a Joint Video Team (JVT), and to create a single technical design called H.264/AVC [1]¹, [2]. The standard text has been approved by ITU-T SG16 as Recommendation H.264 in May 2003 and by ISO/IEC as International Standard 14496-10 (MPEG-4 part 10) Advanced Video Coding (AVC) in July 2003. The primary goals of H.264/AVC are *improved coding efficiency* and *improved network adaptation*. The syntax of H.264/AVC typically permits a significant reduction in bit-rate [3] compared to all previous standards such as ITU-T Rec. H.263 [4] and ISO/IEC JTC 1 MPEG-4 Visual [5] at the same quality level.

The demand for fast and location-independent access to multimedia services offered on today's Internet is steadily increasing. Hence, most current and future cellular networks, like GSM-GPRS, UMTS, or CDMA-2000, contain a variety of packet-oriented transmission modes allowing transport of practically any type of IP-based traffic to and from mobile terminals, thus providing users with a simple and flexible transport interface. The third generation partnership project (3GPP) has selected several multimedia codecs for the inclusion into its multimedia specifications [6]. To provide basic video service in the first release of the 3G wireless systems, baseline H.263 as a mandatory codec and the MPEG-4 visual simple profile as an optional codec have been integrated. The choice was based on the manageable complexity of the encoding and decoding process as well as on the maturity and simplicity of the design.

However, due to the likely business models in emerging wireless systems in which the end-user's costs are proportional to the transmitted data volume and also due to limited resources bandwidth and transmission power, compression efficiency is the main target for wireless video and multimedia applications. This makes H.264/AVC coding an attractive candidate for all wireless applications including Multimedia Messaging Services (MMS), packet-switched streaming services (PSS) and conversational applications.

For efficient transmission in different environments not only coding efficiency is relevant, but also the seamless and easy integration of the coded video into all current and possible future protocol and multiplex architectures. Therefore, H.264/AVC distinguishes between two different conceptual layers, the Video Coding Layer (VCL), and the Network Abstraction Layer (NAL). Both the VCL and the NAL are part of the H.264/AVC standard. The VCL specifies an efficient representation for the coded video signal. The NAL defines the interface between the video codec itself and the outside world. It operates on NAL units which give support for the packet-based approach of most existing networks. The exact transport and encapsulation of NAL units for different transport systems, such as RTP/IP [7], are outside the scope of the H.264/AVC standardization. In addition, for conversational applications the video codec's support of enhanced error resilience features is of major importance. This has also been taken into account in the standardization of this codec.

This paper is structured as follows. Section 2 provides an overview over the H.264/AVC video coding standard from the perspective of wireless video applications. We categorize features according to their applicability in different video services. Section 3 provides experimental results for selected system concepts based on the common test conditions.

¹All referenced standard documents can be accessed via anonymous ftp at ftp://standard.pictel.com/video_site, <ftp://ftp.imtc-files.org/jvt-experts>, <ftp://ftp.ietf.org/>, or <ftp://www.3gpp.org/Specs/archive>.

2 H.264/AVC – An Efficient and Flexible Video Coding Toolbox

2.1 Compression Efficiency and Encoder Flexibility

Although the design of the H.264/AVC VCL basically follows the design of prior video coding standards such as MPEG-2, H.263, and MPEG-4, it contains many new features that enable it to achieve a significant improvement in terms of compression efficiency. We will briefly highlight those. For more details we refer to [1], [2]. A typical encoder with the main encoding options is shown in Figure 1.

The encoding operation for a picture is summarized as follows. The picture to be encoded is split into blocks. The first picture of a sequence or a random access point is typically coded in Intra mode, i.e., without using other information than the information contained in the picture itself. Each sample of a block in such an Intra frame is predicted using spatially neighboring samples of previously coded blocks. The encoding process is to choose which and how neighboring samples are used for Intra prediction which is simultaneously conducted at encoder and decoder using the transmitted Intra prediction side information.

For all remaining pictures of a sequence or between random access points, typically Inter coding is utilized. Inter coding employs prediction (motion compensation) from other previously decoded pictures. The encoding process for Inter prediction (motion estimation) consists of choosing motion data comprising the reference picture and a spatial displacement that is applied to all samples of the block. The motion data which are transmitted as side information are used by encoder and decoder to simultaneously provide the inter prediction signal.

The residual of the prediction (either Intra or Inter) which is the difference between the original and the predicted block is transformed. The transform coefficients are scaled and quantized. The quantized transform coefficients are entropy coded and transmitted together with the side information for either Intra-frame or Inter-frame prediction.

The encoder contains the decoder to conduct prediction for the next blocks or next picture. Therefore, the quantized transform coefficients are inverse scaled and inverse transformed in the same way as at the decoder side resulting in the decoded prediction residual. The decoded prediction residual is added to the prediction. The result of that addition is fed into a deblocking filter which provides the decoded video as its output.

The quantized transform coefficients as well as all other syntax elements of a MB are conveyed by one of two supported entropy coding methods, Context-Adaptive Variable Length Coding (CAVLC) and Context-Adaptive Binary Arithmetic Coding (CABAC). The usage of adaptive codes permits the adjustment to non-stationary symbol statistic and context modeling allows for exploiting statistical dependencies between symbols.

For removing block-edge artifacts, the H.264/AVC coding design includes an in-loop deblocking filter. The H.264/AVC deblocking filter is applied inside the motion prediction loop. The filtering strength is adaptively controlled by the values of several syntax elements.

The normative part of a video coding standard in general only consists of the appropriate definition of the order and semantics of the syntax elements and the decoding of error-

free bit-streams. This allows a significant flexibility at the encoder, which can on the one hand be exploited for pure compression efficiency, on the other hand several included features in the standard can be selected by the encoder for other purposes such as error resilience, random access, etc. These additional features will be categorized to different wireless applications and will be discussed in more details in the following.

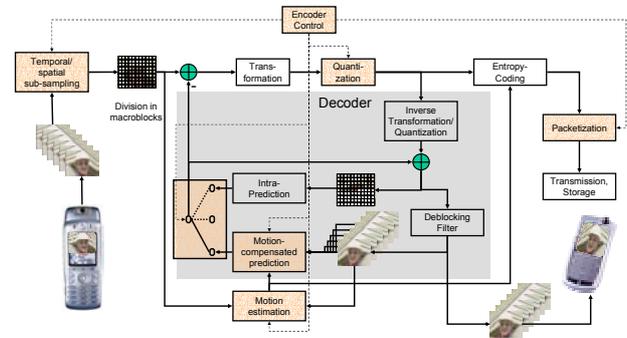


Figure 1 H.264/AVC Encoder realization with coding options.

2.2 Features for Multimedia Messaging and Wireless Packet-based Streaming

For MMS applications, because of the strict separation of encoding, transmission, and decoding, the main issue is compression efficiency. Other helpful features included in the encoding include the insertion of regular intra frames with Instantaneous Decoder Refresh (IDR) for random access and fast forward. The rate control is typically applied such that video quality is almost constant over the sequence, regardless of the scene complexity except for constraints from the Hypothetical Reference Decoder (HRD) [10]. On the wireless link layer reliable transmission strategies including retransmissions are used.

Streaming applications involve more technical challenges than MMS due to online transmission and decoding. In this case, pre-encoded data is requested by the user, which inherently does not allow an adaptation to the transmission conditions such as bit-rate or error rate in the encoding process. However, the receiver usually buffers the received data and starts play-back after a few seconds. Once starting playback, a continuous presentation of the sequence should be guaranteed. Wireless channels commonly provide a constant bit-rate and reliable transmission by using an acknowledged mode within a window of a few seconds. With an appropriate setting of the initial delay and receiver buffer a certain quality of service can be guaranteed [11]. However, long-term variances in the bit-rate due to distance, shadowing, or varying multiuser topology in the supported cell with re-newed resource allocation require channel-adaptive streaming technologies, which allow reacting to variable bit-rate channels. According to [12], these techniques can be grouped into three different categories. *Adaptive media playout* [13] allows a streaming media client, without the involvement of the server, to control the rate at which data is consumed by the playout process. A second technology for a streaming media system is proposed, which makes decisions that govern how to allocate transmission resources among packets. Recent work [14] provides a flexible framework to allow *rate-distortion optimized packet scheduling*. This can be supported, if media streams are pre-encoded with appropriate packet dependencies, possi-

bly adapted to the channel (*channel-adaptive packet dependency control*) [15].

The latter techniques are supported by H.264/AVC by various means. As the streaming server is in general aware of the current bit-rate on the wireless bearer, the transmitter can decide to send one of several pre-encoded versions matched to the channel bit-rate. If the channel rate fluctuates only in a small range, frame dropping of non-reference frames is sufficient - resulting in well-known temporal scalability. Switching of versions can be applied at I frames that are also indicated as instantaneous decoder refresh (IDR) pictures to compensate large scale variations of the channel rate. In addition, H.264/AVC supports efficient version switching with the introduction of switching predictive (SP) pictures. For more details on SP pictures see [17]. Note that quality scalable video coding methods such as MPEG-4 fine-grain scalability (FGS) [16] are not supported by H.264/AVC and such extensions of H.264/AVC are currently not planned.

2.3 Features for Wireless Conversational Services – Rate Control and Error Resilience

A necessary requirement for conversational services is a low end-to-end delay being less than 250 ms. This delay constraint has two main impacts on the video transmitted over wireless bearer services with constant bit-rate. To ensure low delay, fast quantization parameter adaptation is provided to adapt the video bit-rate over a short window. Low delay also only allows temporally backward references in motion compensation, since prediction from future frames would introduce additional delay. Furthermore, the required low delay usually does not allow the wireless bearer to provide a reliable data service due to short-term fading. Therefore, an error-resilient video coding standard suitable for conversational wireless services has to provide error resilience features to combat two problems:

1. it is necessary to minimize the visual effect of errors within one frame, and
2. as errors cannot be avoided, the well-known problem of spatio-temporal error propagation in hybrid video coding has to be limited.

In the following we will discuss all error-resilience features included in the H.264/AVC standard. Packet loss probability and the visual degradation from packet losses can be reduced by introducing slice-structured coding. A slice is a sequence of MBs and provides spatially distinct resynchronization points within the video data for a single frame. No intra-frame prediction takes place across slice boundaries. With that, packet loss probability can be reduced, if slices and, therefore, transmission packets are relatively small, since the probability of a bit-error hitting a short packet is generally lower than for large packets. Moreover, short packets reduce the amount of lost information and, hence, the error is limited and error concealment methods can be applied successfully. However, the loss of intra-frame prediction and the increased overhead associated with decreasing slice sizes adversely affect coding performance and requires additional overhead per slice. Especially for mobile transmission, where the packet size clearly affects loss probability a careful selection of the packet size is necessary. H.264/AVC specifies several enhanced concepts to reduce the artifacts caused by packet losses within one frame. Slices can be grouped by the use of aggregation packets into one packet and, therefore, concepts such as Group-of-Block (GOB) and *Slice Interleaving* [18], [19] are

possible. This does not reduce the coding overhead in the VCL, but the costly RTP overhead of up to 40 bytes per packet can be avoided.

A more advanced and generalized concept is provided by a feature that has been called by the proponents *Flexible MB Ordering* (FMO) [20]. FMO permits the specification of different patterns for the mapping of MBs. FMO is especially powerful in conjunction with appropriate error concealment when the samples of a missing slice are surrounded by many samples of correctly decoded slices. Another error resilience feature 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. For more details on FMO and data partitioning, see [8].

Despite all these techniques, packet losses 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 and quick recovery can only 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 MBs 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 MBs anticipating transmission errors.

2.4 Rate-Distortion Optimized Mode Selection

The provided flexibility requires sophisticated mode selection procedures to exploit the potential of H.264/AVC. The concept of selecting appropriate coding options in optimized encoder designs for many video coding standards is based on rate-distortion optimization algorithms [21][22]. The two cost terms “rate” and “distortion” are linearly combined and the mode is selected such that the total cost is minimized. The Lagrange parameter for appropriate weighting of rate and distortion has to be selected appropriately. In the H.264/AVC test model, the Lagrangian mode selection is used for motion vector search as well as MB mode and reference frame selection.

As already discussed, the tools for increased error resilience, in particular those to limit error propagation, do not significantly differ from those used for compression efficiency. Features like multi-frame prediction or intra MB coding are not exclusively error resilience tools. This means that bad decisions at the encoder can lead to poor results in coding efficiency or error resiliency or both. The selection of the coding mode for compression efficiency can be modified taking into account the influence of the random lossy channel. In this case the encoding distortion is replaced with the expected decoder distortion. For the computation of the expected distortion we refer to, e.g. [23] and 0.

Due to the bi-directional nature of conversational applications it is common that the encoder has knowledge of the experienced channel at the decoder, usually with a small delay. In our framework this can be expressed by the knowl-

edge of a d -frame delayed version of the random channel at the encoder. This characteristic can be conveyed from the decoder to the encoder by acknowledging correctly received slices (ACK), sending a not-acknowledge message (NAK) for missing slices or both types of messages. Though re-transmissions are not feasible in a low-delay environment, the observed channel characteristic are still useful at the encoder even if the erroneous frame has already been decoded and concealed. A technique addressing the problem of continuing error propagation has been introduced, among others, in [24], [25], [26], and [27] under the acronym NEWPRED. The flexibility provided in H.264/AVC to select the MB mode and reference frames on MB or sub-MB basis allows incorporating NEWPRED in a straight-forward manner, different schemes are discussed in 0. In the following we apply a mode in which the reference frames in the encoders multi-frame frame buffer are updated with reception of each ACK and NAK by utilizing the identical error concealment as the decoder applies.

3 Selected Simulation Results

3.1 Compression Efficiency

The compression efficiency features as presented in subsection 2.1 in combination with an appropriate rate-distortion optimized mode selection according to subsection 2.4 permit a significant performance gain when compared to state-of-the-art standards such as MPEG-2, H.263, or MPEG-4. Figure 2 shows Coding performance of H.264/AVC codec compared to state-of-the-art video coding standards for QCIF test sequence foreman at frame rate 10 Hz. Bit-rate savings between 30-50% when compared to H.263 or MPEG-4 are achieved.

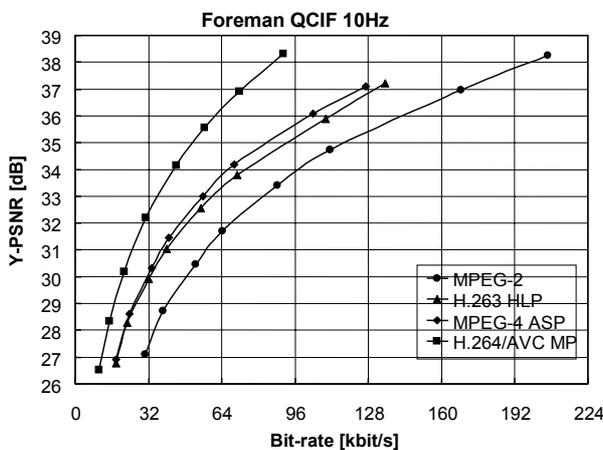


Figure 2 Coding performance of H.264/AVC codec compared to state-of-the-art video coding standards for QCIF test sequence foreman at frame rate 10 Hz.

3.2 Slices and Channel-Adaptive Intra Updates

In the following we will present simulation results based on the wireless test conditions [9] that show the influence of the selection of different error resilience features for the quality of the decoded images. The reported PSNR is the arithmetic mean over the decoded luminance PSNR over all frames of the encoded sequence and over 100 transmission and decoding runs. In addition, we present results for the cumulative distribution of the decoded luminance PSNR for each frame, i.e., the likelihood that the PSNR of the frames of the

sequence is smaller than the value on the x-axis. This shows the variance in the decoded quality. The NAL overhead, the RTP/UDP/IP overhead after RoHC, and the link layer overhead is taken into account in the bit-rate constraints according to [9]. For the following simulations we concentrate on test case 5 from [9], which includes the QCIF test sequence “Foreman” (30 Hz, 300 frames) at a constant frame rate of 7.5 Hz for a mobile user at 3 km/h and maximum bit-rate 64 kbit/s as this is the most critical case in terms of error probability. Rate distortion optimized mode selection and motion vector selection have been used. The distortion and the set of coding modes are appropriately selected according to the applied features.

First we investigate the performance of the system in case that the channel statistics are taken into account into the selection of the coding options in the encoder. For this purpose we replace the encoding distortion by the expected decoder distortion assuming a channel producing independent packet losses with probability p . The packet loss rate for this wireless channel is approximately 4%. Figure 3 shows the cumulative distribution of decoded PSNR for different NAL unit erasure rates for the estimation of the expected distortion in the encoder. Obviously the introduction of loss-aware rate-distortion optimization ($p>0$) increases the decoded quality significantly compared to the results with pure encoding distortion ($p=0$). The average PSNR increases by at least 3 dB when compared to the pure R-D optimization. The advantage of the channel-adaptive mode selection is even more evident when looking at the cumulative distribution of the different strategies. Whereas in case of no error-resilience the probability of bad frames (frames below 22 dB) is at an unacceptable ratio of about 30%, for loss-aware coding this is reduced significantly to less than 8%. It is also obvious that if the expected error rate matches the experienced error rate on the channel, the performance is optimal (see $p=4%$). However, it can also be seen that a mismatch in the expected error rate in the encoder does not have significant influence. Therefore, a rough estimation of the expected decoder distortion at the encoder seems to be good enough for a good mode selection.

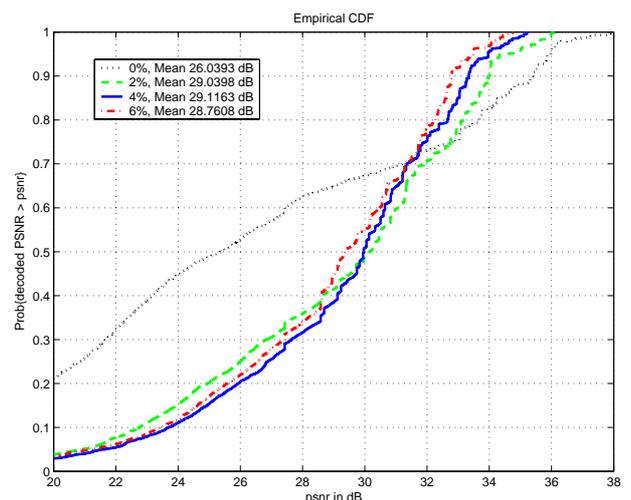


Figure 3 Cumulative distribution of decoded PSNR for different NAL unit erasure rates for the estimation of the expected distortion in the encoder.

The introduction of slices in the encoding has two beneficial aspects when transmitting over wireless channels, but adversely affects the coding efficiency due to increased packet overhead and reduced prediction within one frame, as e.g.

motion vector prediction and spatial intra prediction is not allowed over slice boundaries. The two positive effects with the introduction of slices are the reduced error probability of shorter packets and, the re-synchronization possibility within one frame. The latter technique allows restarting the decoding process at each slice, and, in addition, it allows applying advanced error concealment.

The combination of slice-structured coding and adaptive intra MB updates has been investigated and a comparison with the best cases for channel-adaptive mode selection and slice structured coding with 10 slices per frame and without adaptive intra updates (compare 0) is provided in Figure 4. For the slice-structured coding with encoding distortion ($p=0\%$) the number of packets is selected as $N_p=10$. For the expected decoder distortion without slice-structuring ($N_p=1$) the adapted loss-rate $p=0\%$ is chosen. Finally, for the combination of slice-structured coding and channel-adaptive intra updates the number of packets per frame is selected as $N_p=10$ and, therefore, the appropriate loss probability to compute the expected decoder distortion is about $p=1\%$ (compare 0).

The average decoded PSNR indicates that an optimized combination of both error resilience schemes outperforms each of the presented error resilience schemes significantly. From the cumulative distribution it can be observed that the probability for bad frames below 22 dB in PSNR is almost vanishing for the combined mode.

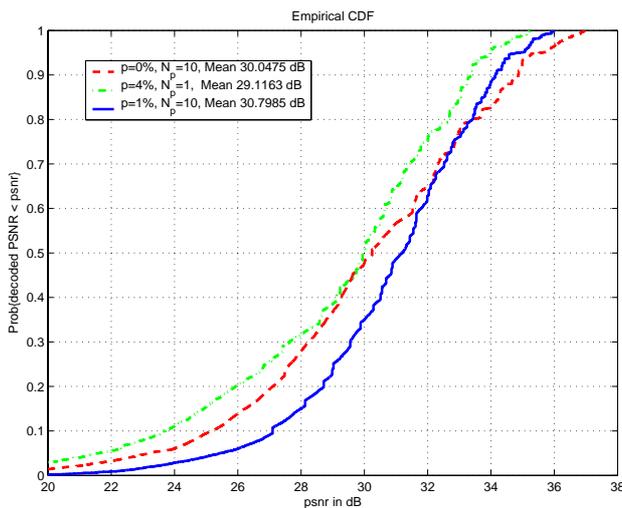


Figure 4 Cumulative distribution of decoded PSNR for different error-resilience strategies: channel-optimized intra updates with and without slice structuring for different assumed loss probabilities p .

3.3 Exploiting Feedback in Video Encoding

Finally, we investigate a system which exploits multiple reference frames and network feedback. We restrict our simulation results to the discussed feedback mode with advanced error concealment. In contrast to the previous simulations we use five reference frames for the feedback mode. Figure 5 shows the cumulative distribution of decoded PSNR for different error-resilience strategies: RD-optimized intra updates with slice structuring, as well as the applied feedback mode with and without slice structuring for delay $d=2$ and $d=4$ frames, which corresponds to a round trip time of about 250 ms and 500 ms, respectively. Let us focus on the delay $d=2$ case first. The results indicate that the optimized intra mode with slice structuring and MB mode

selection with expected decoder distortion and feedback mode 2 perform very similar based on the cumulative distribution and the average decoded PSNR. The feedback mode might still be beneficial in this case as the complex estimation of the decoder distortion is not necessary for the feedback case. However, much more interesting is the case with feedback and no slice structuring. In contrast to the case without feedback (see Figure 5) the renouncement of slice-structured coding provides a significantly higher average decoded PSNR. Initially, this is obviously surprising as packet loss rate is still much lower when several slices are used and also the visual effects for a decoded frame when losing a single slice should be lower than in case of losing an entire frame. The first effect can indeed be observed from the cumulative distribution: The probability of bad frames (PSNR below 22 dB) is higher for $N_p=1$ than for $N_p=10$. However, in case of no errors the increased coding efficiency when not using slices provides many frames with significantly higher PSNR than for slice structuring. As we avoid error propagation, the correctly received frames are really error-free, which is not the case if we use intra updates. Therefore, if feedback information is available and several completely lost frames are tolerable, it is better to use no slice structured coding than harming the compression efficiency. For increased feedback delay $d=4$ the curves are shifted to the left compared to feedback delay 2. However, the $d=4$ and $N_p=1$ performs almost as good as the best case without feedback. Therefore, in case of available feedback, this very simple system without considering expected decoder distortion and slice structuring and just relying on multiple reference frames outperforms many highly-sophisticated error resilience schemes as long as the delay of the feedback is reasonable.

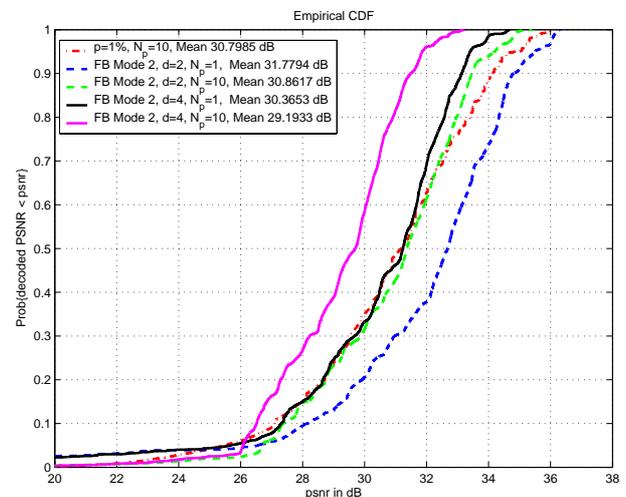


Figure 5 Cumulative distribution of decoded PSNR for different error-resilience strategies: RD-optimized intra updates with slice structuring, and feedback mode with and without slice structuring for delay $d=2$ and $d=4$.

4 Conclusions and Future Work

H.264/AVC promises some significant advances of the state-of-the-art of standardized video coding in mobile applications. In addition to excellent coding efficiency, the design of H.264/AVC also takes into account network adaptation providing large flexibility for its use in wireless applications. The tools provided in H.264/AVC for error resilience do not necessarily differ from the compression efficiency features such as intra MBs or multiple reference frames. However, in

case of error-prone transmission the selection methods have to be changed by using the expected decoder distortion or by restricting the set of accessible coding options. In experimental results based on common test conditions it has been shown that in case without any feedback, several slices in combination with channel-adaptive rate-distortion optimized mode selection is a promising approach. However, in case of available feedback, the application of multiple reference frames to exclude error propagation without slice structuring provides excellent results.

Acknowledgements

The authors would like to thank Stephan Wenger and Miska M. Hannuksela for useful discussions as well as the Joint Video Team (JVT) for the collaborative work and the technically outstanding discussions and contributions.

References

- [1] T. Wiegand (ed.), "Final Committee Draft: Editor's Proposed Revisions," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G50d1, May 2003.
- [2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems*, vol. 13, no. 7, July 2003.
- [3] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-Constrained Coder Control and Comparison of Video Coding Standards," *IEEE Transactions on Circuits and Systems*, vol. 13, no. 7, July 2003.
- [4] ITU-T, "Video coding for low bitrate communication," ITU-T Recommendation H.263; version 1, Nov. 1995; version 2, Jan. 1998.
- [5] ISO/IEC JTC1, "Coding of audio-visual objects – Part 2: Visual," ISO/IEC 14496-2 (MPEG-4 visual version 1), April 1999; Amendment 1 (version 2), February, 2000; Amendment 4 (streaming profile), January, 2001.
- [6] 3GPP Technical Specification 3GPP TR 26.140: "Multimedia Messaging Service (MMS); Media formats and codecs."
- [7] S. Wenger, T. Stockhammer, and M.M. Hannuksela, "RTP payload Format for H.264 Video", draft-wenger-avt-rtp-h264-01.txt, Internet Draft, Work in Progress, March 2003.
- [8] S. Wenger, "H.264/AVC over IP," *IEEE Transactions on Circuits and Systems*, vol. 13, no. 7, July 2003.
- [9] G. Roth, R. Sjöberg, G. Liebl, T. Stockhammer, V. Varsa, and M. Karczewicz, "Common Test Conditions for RTP/IP over 3GPP/3GPP2," ITU-T SG16 Doc. VCEG-M77, Austin, TX, USA, Apr. 2001.
- [10] J. Ribas-Corbera, P.A. Chou, and S. Regunathan, "A Generalized Hypothetical Reference Decoder for H.264/AVC", *IEEE Transactions on Circuits and Systems*, vol. 13, no. 7, July 2003.
- [11] H. Jenkac, T. Stockhammer, and G. Kuhn, "Streaming Media in Variable Bit-Rate Environments", in Proc. Packet Video Workshop, Nantes, France, April 2003.
- [12] B. Girod, M. Kalman, Y. J. Liang, and R. Zhang, "Advances in video channel-adaptive streaming," in Proc. ICIP 2002, Rochester, NY, September 2002.
- [13] E.G. Steinbach, N. Färber, and B. Girod, "Adaptive play-out for low latency video streaming," in Proc. ICIP 2001, Thessaloniki, Greece, October 2001.
- [14] P.A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, submitted February 2001, <http://research.microsoft.com/pachou>.
- [15] Y.J. Liang and B. Girod, "Rate-distortion optimized low-latency video streaming using channel-adaptive bitstream assembly," in Proc. ICME2002, Lausanne, Switzerland, August 2002.
- [16] H. Radha, M. van der Schaar, and Y. Chen, "The MPEG-4 Fine-Grained Scalable Video Coding Method for Multimedia Streaming over IP," *IEEE Trans. Multimedia*, vol. 3, no. 1, March 2001.
- [17] M. Karczewicz, and R. Kurçeren, "The SP and SI Frames Design for H.264/AVC", *IEEE Transactions on Circuits and Systems*, vol. 13, no. 7, July 2003.
- [18] S. Wenger and G. Cotê, "Using RFC 2429 and H.263+ at low to medium bit-rates for low-latency applications", in Proc. Packet Video Workshop 1999, New York, NY, April 1999.
- [19] V. Varsa, and M. Karczewicz, "Slice interleaving in compressed video packetization", in Proc. Packet Video Workshop 2000, Forte Village, Italy, May 2000.
- [20] S. Wenger, and M. Horowitz, "Flexible MB Ordering – A New Error Resilience Tool for IP-Based Video", in Proc. IWDC 2002, Capri, Italy, Sept. 2002.
- [21] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra. "Rate-Distortion Optimized Mode Selection for Very Low Bit Rate Video Coding and the Emerging H.263 Standard," *IEEE CSVT*, 6(2):182–190, April 1996.
- [22] G.J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74-90, Nov. 1998.
- [23] 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.
- [24] B. Girod, and N. Färber, "Feedback-Based Error Control for Mobile Video Transmission," *Proceedings of IEEE*, vol. 97, no. 10, pp. 1707-1723, Oct. 1999.
- [25] T. Nakai, and Y. Tomita, "Core experiments on feedback channel operation for H.263+," ITU-T SG15 LBC96-308, Nov. 1996.
- [26] S. Fukunaga, T. Nakai, and H. Inoue, "Error resilient video coding by dynamic replacing of reference pictures," in Proc. IEEE Globecom, vol.3, Nov. 1996.
- [27] Y. Tomita, T. Kimura, and T. Ichikawa, "Error resilient modified inter-frame coding system for limited reference picture memories," in Proc. Picture Coding Symposium, Berlin, Germany, Sept. 1997.
- [28] T. Stockhammer, M.M. Hannuksela, and T. Wiegand, "H.264/AVC in Wireless Environments," *IEEE Transactions on Circuits and Systems*, vol. 13, no. 7, July 2003.