WIRELESS BROADCASTING USING THE SCALABLE EXTENSION OF H.264/AVC

Thomas Schierl, Heiko Schwarz, Detlev Marpe, and Thomas Wiegand

Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI), Image Processing Department
Einsteinufer 37, 10587 Berlin, Germany
{schierl,hschwarz,marpe,wiegand}@hhi.fraunhofer.de

ABSTRACT

For the ongoing standardization activities for IP-based broadcast video and audio streaming services over DVB-H, DMB or 3GPP's MBMS the use of new forward error protection techniques is investigated, since these activities only propose the use of equal error protection schemes. Especially in distributing video hierarchical protection schemes can be used for more efficient content aware protection. Even older video coding standards allow simple quality reducing mechanism like temporal scalability. New scalable video coding schemes like the scalable extensions of H.264/AVC can be used for a further enhancement of the protection schemes. In combination with an unequal erasure protection scheme the efficiency of network delivery can be enhanced significantly. We propose a scheme, which shows benefits compared to single layer stream protection schemes. The scheme allows protection over a larger range of error rates compared to single layer protection while showing "graceful degradation" like behavior.

1. INTRODUCTION

In this paper we will analyze the use of unequal protection schemes for enhancing the video quality over a wider range of error rates compared with single layer protection schemes by using a "graceful degradation" like functionality. The scheme is based on the UXP (Unequal Erasure Protection) which is also investigated within IETF [1]. We propose a scalable solution as depicted in Fig. 1 addressing an UXP scheme in combination with scalable video coding in error prone environments, the idea of UXP in combination with using different quality layers will be investigated within this scenario. Such a protection approach can ideally provide a base quality with higher probability over a wider range of channel errors rates than single layer protection schemes. Thus an UXP scheme can lead to significant quality enhancements in mobile Broadcast environments, where variation in coverage is one of the most crucial problems.

Moreover we propose a packetization scheme for the scalable extensions of H.264 using the "RTP Payload Format for H.264 video" for the IETF UXP profile.

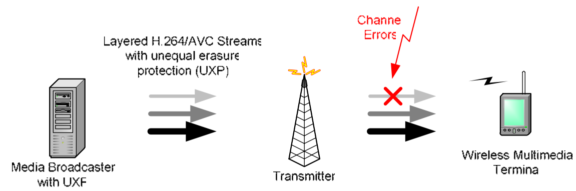

Fig. 1 – Unequal Erasure Protection with scalable H.264/AVC FGS layers

For scalable video coding we will use the scalable extension of H.264/AVC as proposed in [2][3][5], which has been chosen to be the starting point of the new standardization activity on Scalable Video Coding of the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG). The specification of the codec is provided in the first Working Draft [4] for this standardization project and the software can be downloaded at the web site: http://ip.hhi.de/imagecom_G1/savce/index.htm.

2. SCALABLE EXTENSION OF H.264/AVC

The basic idea of the scalable H.264/AVC extension [2][3][4][5] is to extend the hybrid video coding approach of H.264/AVC [6][7] in a way that a wide range of spatio-temporal and quality scalability is achieved. At this, we refer to scalability as a functionality that allows the removal of parts of the bit-stream while achieving a reasonable coding efficiency of the decoded video at reduced temporal, SNR, or spatial resolution. In this paper, we focus on temporal and Fine Grain SNR scalability (FGS).

One key element of the scalable video codec is the hierarchical prediction structure as illustrated in Fig. 2. The first picture of a video sequence is coded as IDR picture; so-called key pictures are coded in regular

0-7803-9332-5/05/$20.00 ©2005 IEEE

intervals. A key picture and all pictures that are temporally located between a key picture and the previous key picture are considered to build a group of pictures (GOP). The key pictures are either intra-coded or inter-coded by using previous key pictures as reference for motion-compensated prediction. The remaining pictures of a GOP are hierarchically predicted as shown in Fig. 2. It is obvious that this hierarchical prediction structure inherently provides temporal scalability; but it turned out that it also offers the possibility to efficiently integrating SNR scalability.

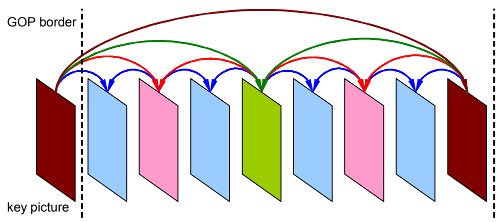

Fig. 2 – Hierarchical prediction structure.

For providing Fine Grain SNR scalability (FGS), a picture is generally represented by an H.264/AVC compatible base representation and zero or more FGS enhancement representations, which represent the residual between the original prediction residuals and intra blocks and their reconstructed base representation. In order to provide quality enhancement layer NAL units that can be truncated at any arbitrary point, the coding order of transform coefficient levels has been modified in a way that the transform coefficient blocks are scanned in several pathes, and in each path only a few coding symbols for a transform coefficient block are coded. For further details, please refer to [2][3][4][5].

Our experiments turned out that the hierarchical prediction structure also increases the coding efficiency. Typically, the maximum coding efficiency is obtained with groups of 16 or 32 pictures (5 or 6 hierarchy stages). The influence on the coding efficiency is even more significant for the FGS enhancement layers. This is related to the fact that in order to enable a re-synchronization between encoder and decoder the FGS enhancements of the key pictures are not used for motion-compensated prediction of following key pictures, and thus the coding efficiency for predicted key pictures gets worse. Thus, a higher coding efficiency for the FGS enhancement layers is generally obtained with larger GOP sizes.

3. UNEQUAL ERASURE PROTECTION

The scalable video coding approach allows the split of the bit-stream into a base layer and various FGS enhancement layers. This allows a transmission of these sub-streams via different channels or in different network streams as well as a transmission of these layers within different protection classes of an UXP transmission profile as proposed in [1]. Forward Error Correction (FEC) schemes based on Reed Solomon codes are often used to protect data sent out via wireless broadcast channels. These codes belong to the class of linear non binary block codes and maintain maximum erasure protection while producing a minimum of redundancy. A (n, k) Reed Solomon (RS) code has a t symbol erasure protection, where n is the code word length in symbols and k the number of information symbols per code word. When using an Galois field GF(2^8) the symbols are represented by bit octets (bytes) and (n=255, n-t) RS code are applied [8]. The UXP profile defines the use of different protection classes in one Transmission Sub Block (TSB) as illustrated in Fig.3.

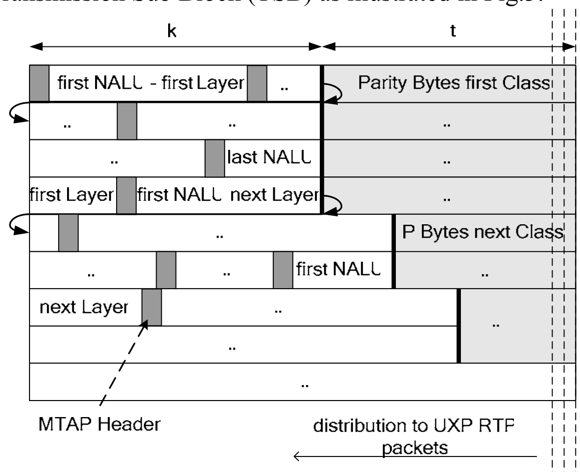

Fig 3 – Scalable H.264/AVC packetization to UXP profile Transmission Sub Block (TSB), in this example the first layer has the highest priority.

A protection class is determined by the erasure capability and thus the numbers of parity bytes. One or more TSBs are placed into one Transmission Block (TB), which is distributed over several RTP packets, where each RTP packet represents one column of the TB as illustrated in Fig. 3. The info stream is inserted into a TSB from the upper left to the lower right, line by line in prioritized order starting with the highest protection class. The smallest unit of an H.264/AVC stream or a stream of the scalable H.264/AVC extension is a so-called Network Abstraction Layer Unit (NALU). In order to avoid an additional rate overhead by filling up each information line with padding bytes, the NAL units of a complete group of pictures (GOP) are directly placed into the TSB in priority order as shown in Fig. 3. The free information bytes for a priority class are filled with the first NAL unit bytes of the next priority class. A reordering to decoding order and timing information assignment is enabled by the use of Multi Time Aggregation Units (MTAP) as proposed in [9] and shown in Fig. 4. MTAP provides the size of the NAL unit, the RTP timestamp offset to the Timestamp carried in the UXP RTP packets and the Decoding Order Number (DON). For the use in the UXP profile the Decoding Order Number Difference (DOND) should be replaced by a an absolute DON value for example of a size of 16 bits for covering the problem of higher interleaving numbers, since no initial MTAP header carrying the DON base number will be present in these packets.

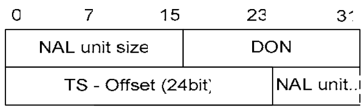

Fig. 4 – modified Multi Time Aggregation Unit (MTAP) header with 16bit DON and 24bit RTP Timestamp Offset.

In this paper, we restrict ourselves to the case of temporal and SNR scalable coding using the FGS concept of the scalable H.264/AVC extension; the spatial scalability feature is not investigated in these first experiments. When we further ignore the SNR scalable representation, the priority order for the NAL units of a GOP is determined by the temporal level of the corresponding pictures. A protection of lower temporal layers that represent the video sequence with a lower frame rate is much more important since due to the motion-compensated prediction and the hierarchical coding structure an error in these pictures influences much more pictures than an error in a higher temporal (enhancement) layer.

The use of FGS will have an additional benefit, when breaking the NAL units from a higher protection class to a lower protection class as shown in Fig. 3, the remaining part in the higher protection class will possibly overcome the channel losses and a cropped FGS NAL unit would remain and enhance the quality additionally.

4 EXPERIMENTAL RESULTS

In the following we analyze the use of different protection schemes for scalable H.264/AVC streams with one base and two FGS enhancement layers with different GOP sizes of 4, 8 and 16 pictures compared to a single layer protection scheme, see Tab. 1. The single layer protection scheme is emulated by using the same protection for each FGS layer of the stream. If one base or FGS packet gets lost, the complete picture is marked as lost. At the moment a single layer stream consisting of a pure H.264/AVC base layer would have a higher coding efficiency compared to the SNR scaled stream with FGS enhancements. The optimization of the FGS functionality is future work and this paper mainly focuses on the influence of different protection schemes.

The test stream consists of a repeated part of the Forman sequence making up 6400 pictures for gathering statistics, see Tab.1. An intra picture is inserted every 2.13 sec.

Pictures in GOP		4	8	16
Resulting Temporal Levels		3	4	5
Base Layer (H264)	**Bitrate / kBit/s**	132.44	124.70	125.15
	PSNR / dB	31.90	32.16	32.58
FGS 0 Layer	**Bitrate / kBit/s**	341.80	255.54	221.12
	PSNR / dB	34.78	34.77	34.94
FGS 1 Layer	**Bitrate / kBit/s**	449.88	449.88	449.88
	PSNR / dB	35.32	36.38	37.06

Tab. 1 – Scalable H.264/AVC test streams with 3 different GOP sizes, 352x288 pels, 30 fps.

We searched for an optimally working protection scheme for the single layer simulation. The condition was a resulting NAL unit loss rate of significantly below 1% at a maximum network packet loss rate of 7.5%. Each picture is contained in a separate NAL unit.

The resulting overhead for that is 20% in transmission rate with the selected protection scheme. This protection scheme treats the temporal level in prioritized way, since lower temporal levels are carrying more information for more pictures. The variation of the fraction of the protection data rate for the different schemes is shown in Tab. 2. For the losses a random loss rate is assumed, that does not exactly represent the behavior of certain broadcast channel, but should be sufficient for testing the protection schemes in general.

Scheme:	Layer:	Base	FGS 0	FGS 1
Single		100%	100%	100%
FGS 0 only		+160%	-100%	-100%
Combined		+23%	+10%	-19%

Tab. 2 – Avg. protection data rate fractions for the base and FGS layer compared to the single protection scheme, as an example for the GOP16 stream. Single protection is treated as one layer.

In our simulation we set the maximum packet size to 1400 bytes, thus multiple GOPs are placed into the UXP RTP packets belonging to one TB. The packetization includes the MTAP header generation. In case of picture loss the last picture is repeated in order to allow a reasonable PSNR calculation.

The results are shown in Fig. 5 – Fig. 7 for GOP sizes of 4, 8 and 16, respectively. The curve of the single layer protection scheme shows with all 3 GOP sizes a good performance up to 10% network packet loss. At this point the resulting packet loss causes a decrease of 0.4 dB for a GOP size of 16 and 0.8 dB for a GOP size of 4 respectively. Below that loss rate the “single layer” scheme is working very well. Above that loss rate the scheme is breaking down fast and looses significantly many dBs with increased network packet loss rate.

Comparing the “FGS0 Layer only” protection scheme it can be noticed, that small loss rates below 5% will destroy the FGS enhancement layers, since they are not protected. The result is a strong protection for the base layer, which is working up to 30% network packet loss and more.

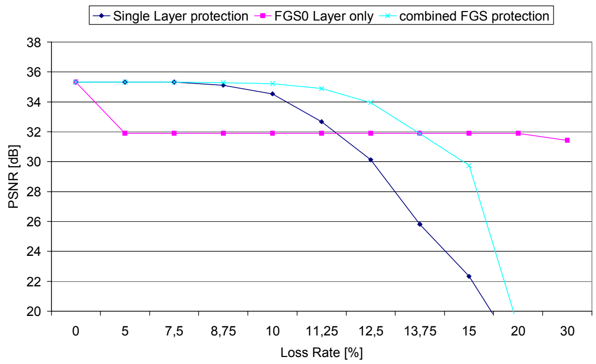

Fig. 5 – Quality over network loss rate, GOP size 4.

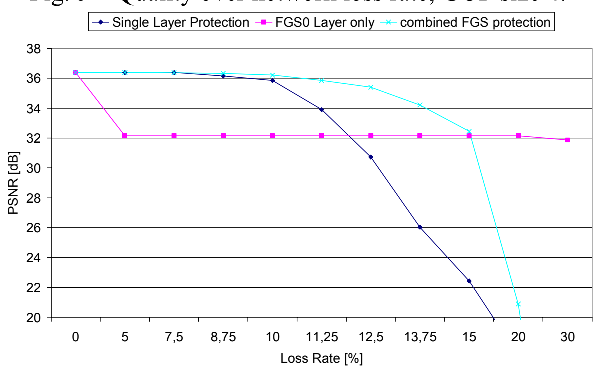

Fig. 6 – Quality over network loss rate, GOP size 8.

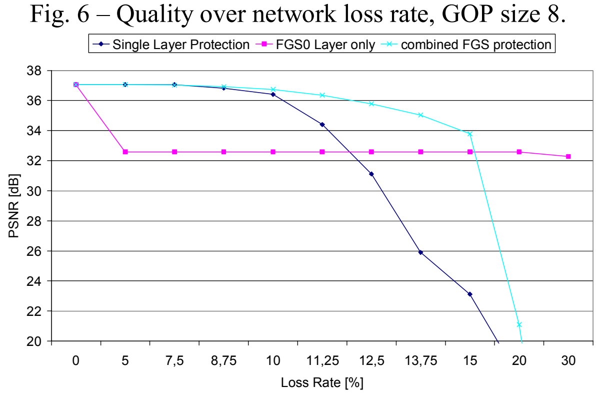

Fig. 7 – Quality over network loss rate, GOP size 16

Comparing the “combined FGS protection” scheme with the “Single layer”, this scheme is outperforming the “Single layer” variation. The “combined FGS protection” works up to 12.5% network packet loss and shows here a loss of about 1 dB. With an increasing loss rate this scheme is not breaking down. Up to a loss rate of 13.75% with a GOP size of 4 and up to 15% with a GOP size of 8 and with a GOP size of 16 a PSNR above 32 dB can be reached. Analyzing the different GOP sizes of the schemes shows that with increased GOP size the “Single layer” as well as the “combined FGS” protection scheme work better. This effect is combined with the higher coding efficiency of streams with higher GOP sizes. But with increased GOP size the “combined FGS” protection scheme significantly gains compared to the “Single layer” scheme.

With increasing network packet loss rate the “combined FGS protection” scheme has the best performance while maintaining a “graceful degradation” like behavior.

5 CONCLUSION

We have presented an approach for using the scalable extension of H.264/AVC with an unequal erasure protection scheme. We have shown that a protection scheme optimized for a quality scalable stream outperforms a single layer protection scheme providing a “graceful degradation” behavior. The increase of the GOP size is improving the gain of a FGS layer prioritized protection compared to a single layer protection scheme.

The optimization of a protection scheme for a certain scalable H.264/AVC stream is future work. Further a stronger use of the FGS behavior in the assignment of data fractions to UXP protection classes shall be investigated in future work.

6 REFERENCES

[1] G. Liebl, M. Wagner, J. Pandel, W. Weng, “An RTP Payload Format for Erasure-Resilient Transmission of Progressive Multimedia Streams”,IETF, October 2004.

[2] H. Schwarz, T. Hinz, H. Kirchhoffer, D. Marpe, T. Wiegand, "Technical Description of the HHI proposal for SVC CE1", ISO/IEC JTC1/SC29/WG11, M11244, Oct. 2004.

[3] H. Schwarz, D. Marpe, T. Schierl, T. Wiegand, “Combined Scalability Support for the scalable Extensions of H.264/AVC”, ICME, accepted paper, July 2005

[4] J. Reichel, H. Schwarz, M. Wien (eds.), “Scalable Video Coding – Working Draft 1,” Joint Video Team (JVT), Doc. JVT-N020, Hong Kong, CN, Jan. 2005.

[5] H. Schwarz, D. Marpe, T. Wiegand, "MCTF and Scalability Extension of H.264/AVC," Proc. of PCS, San Francisco, CA, USA, Dec. 2004.

[6] ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, "Advanced Video Coding for Generic Audiovisual Services", version 3: 2005.

[7] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. CSVT, vol. 13, no. 7, pp. 560-576, July 2003.

[8] S. Lin, D. J. Costello, “Error Control Coding: Fundamentals and Applications”, Prentice-Hall, Inc., N.J., 1983.

[9] S. Wenger, M.M. Hannuksela, T. Stockhammer, M. Westerlund, D. Singer, “RTP Payload Format for H.264 Video“, RFC3984, IETF, February 2005.