Sławomir Maćkowiak
Institute of Electronics and Telecommunication
Poznań University of Technology
Piotrowo 3A
60-965 Poznań
tel. 6652071, fax 6652572
e-mail: smack@et.put.poznan.pl

VIDEO SCALABILITY IN WIRELESS NETWORKS

***Abstract*: Scalability became an extremely important functionality of video codecs employed in error-prone communication environments like wireless transmission systems. The paper describes a multi-layer video coder based on spatio-temporal scalability and data partitioning. The proposed coder combines spatial and temporal scalability with FGS (Fine Granularity Scalability). The proposed solution introduces minor modifications of the bitstream semantics and syntax. The coder consists of two independently motion-compensated sub-coders that encode a video sequence and produce two bitstreams corresponding to two different levels of spatial and temporal resolution. The functionality of FGS is related to some drift in the enhancement layer. This drift can be limited by excluding temporal prediction in some enhancement layer frames.**

1. INTRODUCTION

Recently, scalability became an extremely important functionality of video codecs employed in communication networks. Scalability means that a video data bitstream is partitioned into layers in such a way that the base layer is independently decodable into a video sequence with reduced spatial resolution, temporal resolution or signal-to-noise ratio (SNR) [16]. Enhancement layers provide additional data necessary for video reproduction with higher spatial resolution, temporal resolution or signal-to-noise ratio. This functionality is called spatial, temporal or SNR scalability, respectively, as defined by video coding standards: MPEG-2 [1, 2] and MPEG-4 [3, 5].

Scalability provides an opportunity to broadcast data once to a group of users accessed via heterogeneous links characterized by various levels of quality of service, e.g. available bitrate [6]. The importance of scalability is being recognized as more attention is paid to video transmission in error-prone environments, such like wireless video transmission systems [7, 8]. A video bitstream is error sensitive due to extensive employment of variable-length coding causing that a single transmission error may result in an undecodable long string of bits. It was shown that an efficient method of improving transmission error resilience is to split the coded video bitstream into a number of separate bitstreams (layers) transmitted via channels with different degrees of error protection, i.e. the base layer is better protected while the enhancement layers exhibit lower level of protection [9÷12]. Application of scalable coders yield better subjective quality in the case of packet losses and this quality improvement is particularly visible for spatial scalability [9]. Similarly, video bitstream partitioning substantially increases error resilience of compressed video transmitted via wireless links [12, 13]. In wireless networks, scalable video can withstand bandwidth variations and assist rate control during congestions [14].

The state of the art in video compression has just experienced a revolution with the new standard H.264/MPEG-4 AVC/MPEG-4 Part 10. Version 1 of the new video coding standard AVC/H.264 has been already developed by the JVT Committee [10,11]. Many advanced techniques implemented in AVC make it the most efficient and powerful hybrid video coder.

H.264/AVC offers a significant improvement of coding efficiency compared to other compression standards such as MPEG-2. The H.264/MPEG-4 AVC delivers 50 percent improvement of bitrate reduction, representing the most significant technological advancement in coding efficiency and quality since MPEG-2/H.262.

As with MPEG-2, H.264/AVC is based on block transforms and motion compensated predictive coding. H.264/AVC leverages today's processing power to provide improved coding techniques, including multiple reference frames and several block sizes for motion compensation, intra-frame prediction, a new 4x4 integer transform, a 1/4 pixel precision motion compensation, an in-the-loop deblocking filter, and improved entropy coding.

The AVC/H.264 compliant coder is capable of delivering sub 1 Mbps video at DVD quality, which could let operators to offer more channels in existing systems, and let consumers store twice as many programs in personal video recorders or to record high-definition content onto DVD.

AVC/H.264 codec can be used over the full range of video coding available under MPEG from coding video transmitted over cellular networks upto the transmission of high quality high-definition video.

Rapid development of video communication, including wireless video, still yields urgent need to improve flexible bit allocation to individual layers, i.e. fine granularity scalability (FGS) which is also already proposed for MPEG-4 [14, 15] where the fine granular enhancement layers are intraframe encoded, thus reducing coding efficiency. This way of evolution should be expected also for AVC/H.264 video codec with functionality of scalability. This proposal is similar to that one already made for MPEG-2 and H.263, and it extends the ideas already proposed in the context of classic hybrid video coders [12-13]. This paper reports results for modified AVC/H.264 and MPEG-2 video codecs with FGS scalability.

2. PROBLEM STATEMENT

Scalable video coding involves generating a coded representation (bitstream) in a manner that facilitates the derivation of video of more than one resolution or quality from this bitstream. In scalable video coding, the total bitrate is the sum of layer bitrates.

In simulcast coding, each bitstream of video is associated with a certain resolution or quality and is encoded independently. Thus, any bitstream can be decoded by a single-layer decoder. The total bitrate required for transmission of encoded streams is the sum of bitrates of these streams. Therefore, this technique is not efficient.

For a given overall decoded video quality, scalable coding is not acceptable in common applications, if the bitrate is significantly greater than the bitrate achieved in single-layer coding.

Most scalability proposals are based on one type of scalability. The universal scalable coding has to include different types of scalability. A single scalable video technique cannot serve a broad range of bitrates in networks (e.g. from a few kbps to several Mbps) or a wide selection of terminals with different characteristics.

Among various possibilities, the combination of spatial and temporal scalability with SNR scalability seems very promising. It provides a possibility to produce one bitstream which represents an encoded sequence with two different spatial and temporal resolutions at the same time.

The current version of AVC encoder does not support scalability. Due to the fact that such functionality is very important nowadays, it is vital to include it into this new advanced codec.

Coding efficiency of natural hybrid video coders is substantially improved as compared to the simulcast coding.

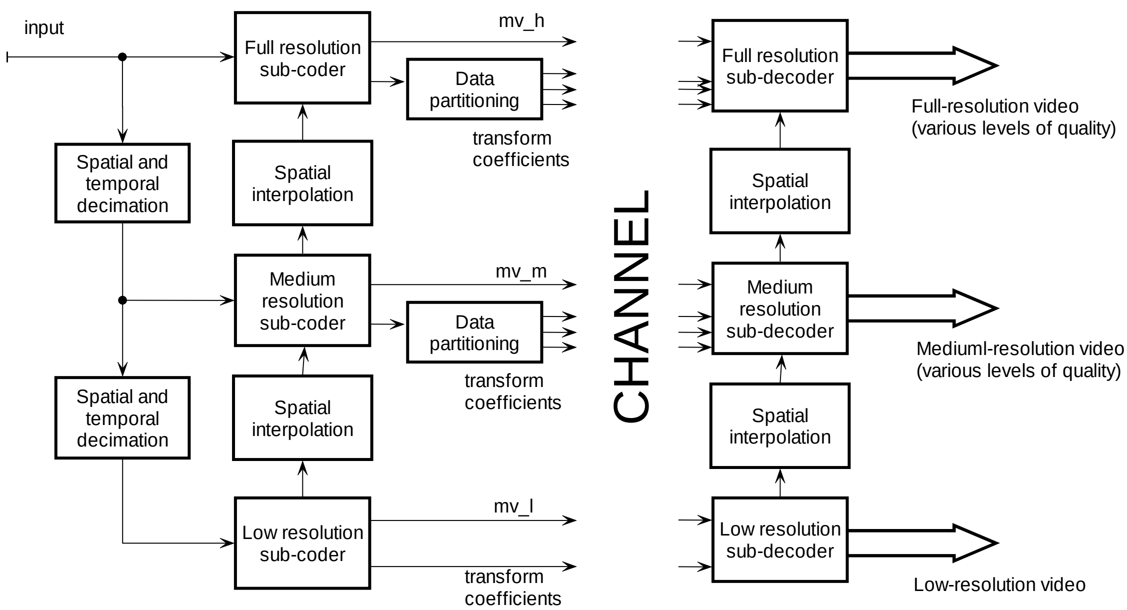


Fig. 1. Multi-layer video coder with patio-temporal and SNR scalability. mv_l , mv_m and mv_h denote motion vectors from the low-resolution, medium-resolution layer and the full-resolution layer, respectively.

The goal of the paper is to describe a scalable extension of the AVC coder. The assumption is to introduce possibly minor modifications of the bitstream semantics and syntax as well as to avoid as much as possible the technologies that are not present in the existing structure of the AVC codec. In particular, it is assumed that the low-resolution base layer bitstream is fully compatible with the AVC/H.264 standard. Moreover, the bitstream syntax is standard, and minor semantics modifications are proposed for the enhancement layer only.

3. GENERAL CODER STRUCTURE

The scalable coder proposed consists of some motion-compensated coders that encode a video sequence and produce bitstreams corresponding to different levels of spatio-temporal resolutions. For example, a three-layer video representation may be produced by a three-loop video coder (Fig. 1).

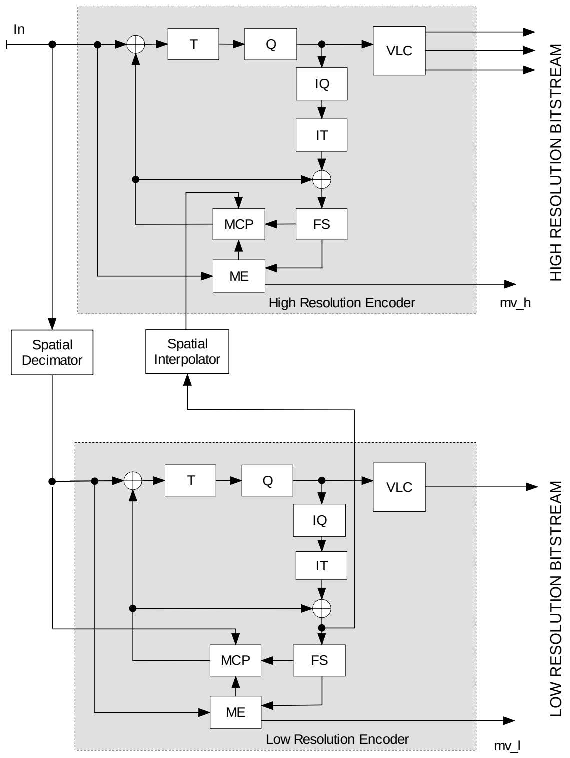
For in-depth analysis, a two-loop encoder has been chosen. In fact, our coder consists of two motion-compensated sub-coders (Fig. 2). Each of the sub-coders has its own prediction loop with independent motion estimation and compensation. Data partitioning is used in order to obtain the FGS functionality. Further detailed considerations and experiments deal with data partitioning in the full-resolution enhancement layer only. For the horizontal, vertical and temporal subsampling factors of 2, the range of bitrate matching due to FGS extends mostly from about 30% to 100% of the total bitrate for a scalable coder.

The low-resolution sub-coder is implemented as a standard motion-compensated hybrid AVC coder that produces a bitstream with fully standard AVC syntax. The high-resolution sub-coder is a modified AVC coder that is able to exploit the interpolated macroblocks from the decoded base-layer bitstream. These interpolated macroblocks are used as reference macroblocks for prediction whenever they provide lower cost. Other additional reference macroblocks are created by averaging the reference of temporal prediction and the interpolated macroblock.

Good performance of spatio-temporal down- and upsampling is critical for good performance of the whole coding process.

Spatial decimation includes spatial lowpass filtering that prevents spatial aliasing in the base-layer low-resolution sequence. The choice of the filter trades off between high aliasing attenuation and short temporal response. The results of experimental comparisons prove the importance of the careful choice of the decimation-interpolation scheme.

The system considered employs edge-adaptive bi-cubic interpolation as described in [15]. The technique is applicable to both luminance and chrominance.



T	Transform (e.g. DCT)
Q	Quantization
IQ	Dequantization
IT	Inverse transform
FS	Frame memory
MCP	Motion-compensated predictor
ME	Motion estimator

Fig.2. The structure of the encoder (temporal subsampling is not included in this figure). VLC – variable-length coder. mv_l and mv_h denote motion vectors from the low-resolution and the full-resolution layer, respectively.

Temporal scalability is achieved using bi-directionally predicted frames, or B-frames.

B-frames are disposable, since they are not used as reference frames for the prediction of any other frames. This property allows B-frames to be discarded without destroying the ability of the decoder to decode the sequence and without adversely affecting the quality of any subsequent frames, thus providing temporal scalability.

In this paper, temporal resolution down-sampling is achieved by partitioning the stream of B-frames: every second frame is skipped in the low resolution layer.

4. EXPERIMENTAL RESULTS

The experiments were conducted to test the efficiency of allocation of DCT coefficients between layers. The experimental results were prepared for new AVC/H.264 encoder and the classic MPEG-2 scalable encoder.

4.1. AVC/H.264 WITH FGS SCALABILITY

The scalable test model has been implemented on the standard JVT software version 2.1.

In order to test the coding performance of the scalable AVC codec a series of experiments have been performed with (352 x 288)-pixel sequences. Horizontal, vertical and temporal subsampling factors have been set to 2 and the video sequence structure was that from Fig. 3.

In the experiments, the following modes have been switched on:

- CABAC coder,
- ¼-pel motion estimation in both layers,
- all prediction modes.

Fine granularity scalability (FGS) is obtained via data partitioning in the UVLC (exp-Golomb) coding mode. Except from headers and motion vectors, the bitstreams can be arbitrarily split into layers and multi-layer fine granularity can be achieved (Fig.3).

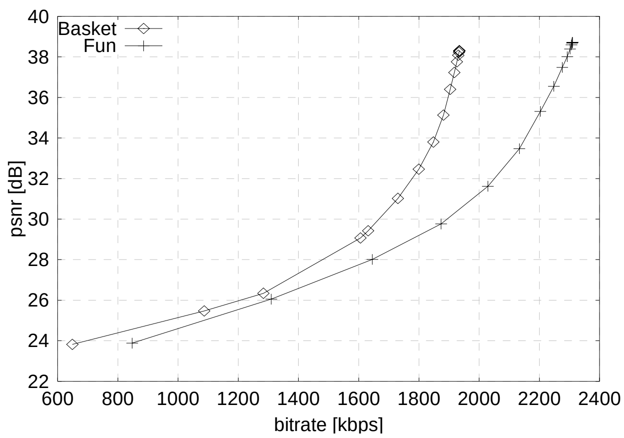


Fig. 3. Rate-distortion curves for FGS in the extended AVC codec: test sequences Funfair and Basket.

4.1. MPEG-2 WITH FGS SCALABILITY

In order to evaluate the proposed partitioning of DCT coefficients in MPEG-2 scalable encode, a verification model has been written in the C++ language and is currently available for progressive 4CIF (704 x 576), 50 Hz, 4:2:0 video test sequences. This software also provides an implementation of the MPEG-2 encoder, which has been cross-checked with the MPEG-2 verification model. The experiments fulfilled the following conditions:

- motion estimation with half-pixel accuracy,
- full search of motion vectors in range [-31,32],
- independent motion estimation and compensation in both layers,
- independent control of the bitrate in the base and enahancement layers,
- the GOP structure of the enhancement layer:
- I–BE–BR–BE–P–BE–BR–BE– P–BE–BR–BE,
- 12 frames in a GOP.

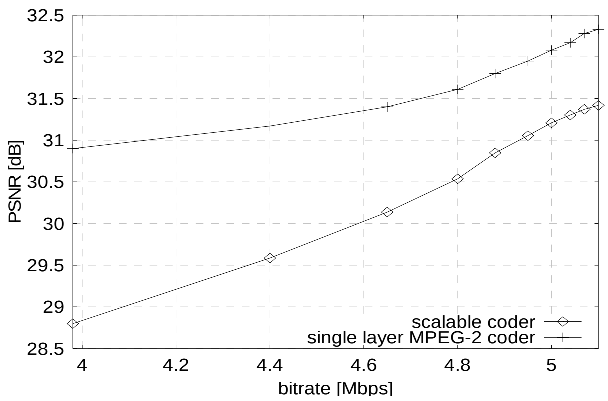


Fig. 4. Fine granularity scalability in a two-loop encoder (lower curve) compared to a single layer MPEG-2 encoder (upper curve). Test sequence Funfair, total bitrate 5 Mbps, base layer bitrate about 1.66 Mbps, length of GOP=12.

The drawback of this strategy is accumulation of drift. Drift is generated by partitioning the full resolution bitstream. Moreover, when the enhancement layer bitstream is corrupted by errors during transmission, the enhancement layer DCT coefficients cannot be properly reconstructed due to the loss of DCT information. This causes drift between the local decoder and remote decoder. It means that the decoding process exploits only the base layer bitstream.

In some applications, drift is not a significant problem. Drift propagation is limited by insertion of I-frames into the enhancement layer. Such additional enhancement-layer I-frames are encoded using less numbers of bits than single-layer I-frames. It is because the bitstream syntax of these frames is that of P-frames but with no motion vectors and with the interpolated base-layer frames used as reference frames. Furthermore, drift in the full resolution part may be reduced by more extensive use of the low resolution images as reference.

5. CONCLUSIONS

In this paper the author has proposed and discussed a multi-layer system with fine granularity, based on a slightly modified structure of the classic MPEG-2 scalable encoder and new AVC/H.264 encoder. This solution is characterized by flexible bit allocation and low bitstream overhead. The total bitstream increases by about 3% per each layer when data partitioning is used.

REFERENCES

[1] [1]ISO/IEC IS 13818-2 / ITU-T Rec. H.262, "Generic coding of moving pictures and associated audio, part 2: video."

[2] B.G. Haskell, A. Puri, A.N. Netravali, Digital video: an introduction to MPEG-2, New York, Chapman & Hall, 1996.

[3] ISO/IEC IS 14496-2, "Generic coding of audio-visual objects, Part 2: Visual."

[4] ISO/IEC JTC1/SC29/WG11/Doc. N3908, "MPEG-4 Video Verification Model version 18.0," January 2001.

[5] Tutorial issue on the MPEG-4 standard, Signal Processing: Image Communication, vol. 15, Nos. 4-5, January 2000.

[6] W. Tan, A. Zakhor "Video multicast using layered FEC and scalable compression," IEEE Trans. Circ. Syst. Video Techn., vol. 11, pp. 373-385, March 2001.

[7] Special issue on video transmission for mobile multimedia applications, Proc. of the IEEE, vol. 87, October 1999.

[8] L. Hanzo, P. Cherriman, J. Streit, Wireless video communications, New York, IEEE Press 2001.

[9] R. Stedman, H. Gharavi, L. Hanzo, and R. Steele, "Transmission of subband coded images via mobile channels," IEEE Trans. Circuits Syst. Video Technol., vol. 3, pp. 15 26, Feb. 1993.

[10] ISO/IEC/SC29/WG11/MPEG02/N4920, ISO /IEC 14496-10 AVC | ITU-T Rec. H.264, Text of Final Committee Draft of Joint Video Specification, Klagenfurt, July 2002.

[11] IEEE Transactions on Circuits and Systems for Video Technology Special Issue on the H.264/AVC Video Coding Standard July 2003 Volume 13 Number 7 p.557 - p.726.

[12] H. Gharavi and C.I. Richards, "Partitioning of MPEG coded video bitstreams for wireless transmission," IEEE Signal Processing Letters, vol. 4, pp. 153 155, June 1997.

[13] H. Gharavi and S. Alamouti, "Multipriority video transmission for third-generation wireless communication systems, Proc. of the IEEE, vol. 87, pp. 1751-1763, October 1999.

[14] D. Wu, Y. Hou, Y.-Q. Zhang, "Scalable video coding and transport over broad-band wireless networks," Proc. of the IEEE, vol. 89, pp. 6-20, January 2001.

[15] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," IEEE Trans. Circ. Syst. Video Techn., vol. 11, pp. 301-317, March 2001.

[16] S. Maćkowiak, "Scalable Coding of Digital Video", Doctoral dissertation, Poznań University of Technology, Poznań 2002.