SPATIO-TEMPORAL SCALABILITY USING MODIFIED MPEG-2 PREDICTIVE VIDEO CODING

Adam Łuczak, Sławomir Maćkowiak, Marek Domański

Poznań University of Technology, Institute of Electronics and Telecommunications,
Piotrowo 3A, 60-965, Poznań, POLAND,

Tel: +48 61 66 52 171, +48 61 66 52 762,; fax: +48 61 66 52 572
e-mail: {aluczak, smack, domanski}@et.put.poznan.pl

ABSTRACT

The paper describes spatio-temporally scalable MPEG video coders proposed. Such an encoder produces two bitstreams: base layer bitstream which represents a video sequence with low spatial and temporal resolution and an enhancement layer bitstream which provides additional data needed for reproduction of pictures with full resolution and full temporal frequency. The bitrate overhead measured relative to the single layer MPEG-2 bitstream varies between 2% and 22% for progressive television test sequences. The base layer bitstream constitutes 34-40% of the overall bitstream. The base layer encoder is fully compatible with the MPEG-2 video coding standard. The enhancement layer encoder is a modified version of that used by MPEG-2 for spatial scalability.

1 INTRODUCTION

Rapid development of multimedia services is stimulating interests in video transmission through heterogeneous communication networks that are characterized by various available levels of quality of service. Different levels of quality of service are mostly related to different available transmission bitrates. On the other hand, the service providers demand that the data are broadcasted once to a group of users accessed via heterogeneous links. For this purpose, the transmitted bitstream has to be partitioned into some layers, i.e. the encoder should be scalable. In the simplest case, there are two layers: the base layer and the enhancement layer. The base layer is decodable independently from the enhancement layer and it represents a video sequence with reduced spatial resolution, temporal resolution or signal-to-noise ratio (SNR). The enhancement layer bitstream provides additional data needed for reproduction of pictures with original quality. The respective functionality is called spatial, temporal or SNR scalability.

Similar demands appear for video transmission in error-prone environments where scalable systems have proven to be more robust. The base layer bitstream can be better protected against transmission errors than the whole bitstream.

This paper deals with scalable two-layer systems. Nevertheless the results are extendable also for multi-layer scalable video coding systems.

The paper is focused on functionality of spatial scalability which is already provided in the MPEG-2 [1,2] and MPEG-4 [3] video compression standards. Unfortunately, standard implementations of spatial scalability are mostly related to unacceptably high bitrate overheads as compared to single-layer encoding of video.

The goal of the work is to achieve total bitrate related to both layers of a scalable bitstream possibly close to the bitrate of single-layer coding. The assumption is that high level of compatibility with the MPEG video coding standards would be ensured. In the paper, the MPEG-2 video coding standard is used as reference but the results are also applicable to the MPEG-4 systems with minor modifications. In particular, it is assumed that the low-resolution base layer bitstream is fully compatible with the MPEG-2 standard.

2 SPATIAL SCALABILITY USING SUBBAND DECOMPOSITION

Recently, quit many proposals have been published for more efficient spatially scalable video coding also called as hierarchical video coding. Most of the proposals exploit subband/wavelet decomposition [4-11]. In the case of two-layer systems decomposition into four spatial subbands is used. The subband LL of lowest frequencies constitutes a base layer while the other three subbands are jointly transmitted in an enhancement layer. Unfortunately, in most of such coders, it is difficult to allocate appropriate number of bits to the base layer and to the enhancement layer. First of all, the base layer bitstream (which corresponds to the LL subband) is mostly larger than the enhancement subbands. Because of applications mentioned above, a practical requirement is that the bitstream of the base layer does not exceed the bitstream of the enhancement layer. In order to meet this requirement, it has been proposed to combine the spatial scalability with other types of scalability thus reducing the base layer bitrate. The recent proposals are the following:

- Combination of spatial and SNR scalability [12],
- Combination of spatial and temporal scalability [13, 14, 15].

The latter approach was considered in two versions hitherto. The first one exploited three-dimensional spatio-temporal decomposition [13]. The subband of the lowest spatio-temporal frequencies is transmitted as the base layer while the other subbands correspond to the enhancement layer with the exception of some subbands of high temporal and spatial frequencies which are discarded.

The second approach was based on partitioning of data related to B-frames [14, 15]. This approach as well as that based on combination space and SNR scalabilities were quit successful . Their scalability costs mostly did not exceed 10-20% of total bitrate compared to a non-scalable coding.

3 APPROACH PROPOSED

Here, we also follow the principle of combination of spatial and temporal scalability, i.e. spatio-temporal scalability. In contrary to the previously mentioned references, we are going to show that subband decomposition is not necessary for P- and B-frames and good results are obtainable even by proper exploitation and modification of the prediction schemes already standardized by MPEG-2.

The system proposed exploits the prediction schemes defined in MPEG-2 for spatial scalability [1,2] with temporal scalability algorithms [2,16].

Temporal resolution reduction is achieved by partitioning of the stream of B-frames: each second frame is included into the enhancement layer only. Therefore there exist two types of B-frames: BE-frames which exist in the enhancement layer only and BR-frames which exist both in the base and enhancement layers (Table 1).

In order of simplicity, it is assumed that an odd number of B-frames is located between two consecutive not-B frames (I-frames or P-frames). Therefore the I- and P-frames exist in both layers in the time instants (Table 1).

Table 1. Frames of different types in both layers.

Image type	*Base layer*	*Enhancement layer*
I	Always in both layers	
P	Always in both layers	
BR	Always in both layers	
BE	Does not exist	Enhancement layer only

Table 1 shows that there exist two types of B-frames: BE-frames that exist in the enhancement layer only and BR-frames that exist in both layers. An exemplary but typical GOP structure is as follows:
I–BE–BR–BE–P–BE–BR–BE– P–BE–BR–BE– P–BE–BR–BE
or
I–BE–BR–BE– BR–BE–P–BE–BR–BE– BR–BE– P–BE–BR–BE. .

4 ENCODER STRUCTURE

The encoder will be presented for progressive video.

Base layer coder is implemented as a motion-compensated hybrid MPEG-2 coder. The bitstream produced is described by fully standard syntax. The motion vectors MV for the low-resolution images are estimated independently from those estimated for the enhancement layer. Motion vectors MV are transmitted for the base layer.

The enhancement layer coder is a combination and modification of the MPEG-2 coder with the functionalities of spatial and temporal scalability. In particular, motion is estimated for full-resolution images and full-frame motion compensation is performed. Therefore the number of motion vectors MV_e sent in the enhancement layer is four times that of the base layer.

The overall structure of the coder is shown in Fig. 1.

4.1 P-frame coding

P-frames in the enhancement layer are predicted both from the previous reference frame as well as from the interpolated current low-resolution frame encoded in the base layer. For each macroblock, the best prediction is selected among the two motion-compensated blocks, or an average of them.

4.2 B-frame coding

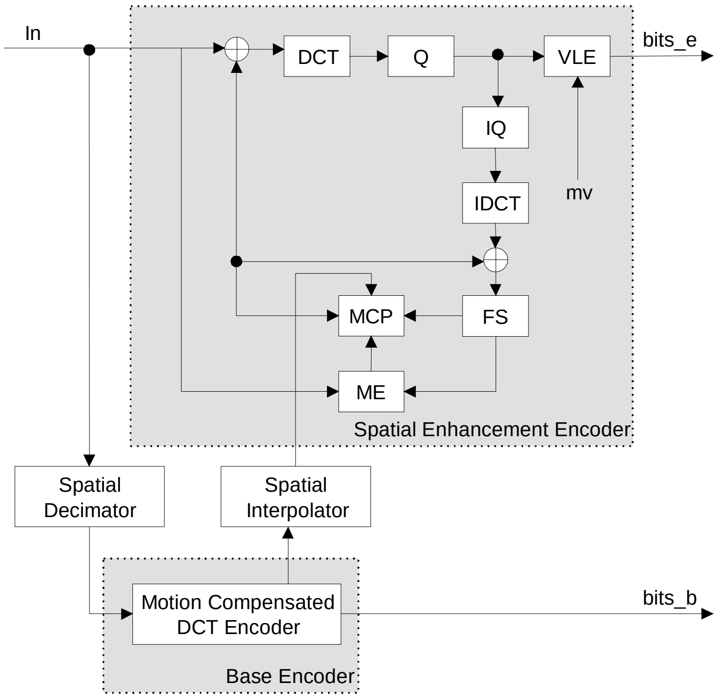
Improved prediction is proposed for the BR-frames, which are the B-frames represented in both layers [17]. Each macroblock in a full-resolution BR-frame can be predicted from the following three reference frames (Fig. 2):

- previous reference frame RP (I- or P-frame from the enhancement layer),
- next reference frame RN (I- or P-frame from the enhancement layer),
- current reference frame RC (BR-frame from the base layer).

Therefore any of the respective reference macroblocks can be used as well as an arbitrary linear combination of two or three of them.

The BE-frames that exist in the enhancement layer only are predicted using a modified procedure from MPEG. The full resolution BR-frames are used as reference frames for BE-frames. Therefore higher correlation between the currently encoded BE-frame and the reference frame is achieved because of a decreased time difference.

The improvement on standard MPEG spatially scalable coding consists in usage of three reference frames (Fig. 3) instead of choosing the best reference from temporal prediction and spatial interpolation. Experimental results with television test sequences prove that this improvement reduces an average size of a BR frame by about 6 - 10% as compared to spatially scalable coding defined in the MPEG-2 standard.



DCT	Discrete cosine transform
Q	Quantization
IQ	Dequantization
IDCT	Inverse cosine transform
FS	Frame memory
MCP	Motion-compensated predictor
ME	Motion estimator

Figure 1. The overall coder structure.

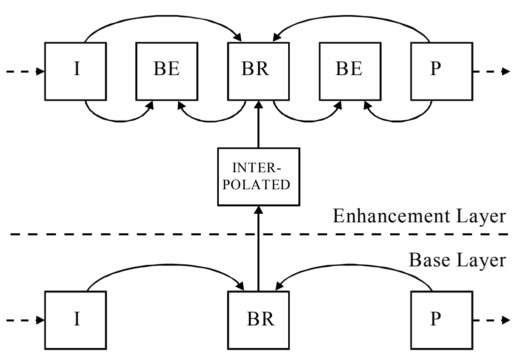


Figure 2. Prediction of B-frames.

4.3 I-frame coding

In the experiments, subband decomposition has been used for I-frame coding only. The subband of lowest frequencies constitutes a base layer while the other three subbands are jointly transmitted in an enhancement layer.

5. EXPERIMENTAL RESULTS

The verification model of the coder proposed has been prepared and tested for progressive 720×576, 50 Hz, 4:2:0 test sequences. The results are given in Table 1. The bitrate overhead measured relative to the single layer MPEG-2 bitstream varies between 2% and 22% for progressive television test sequences. The base layer bitstream constitutes 34-40% of the overall bitstream.

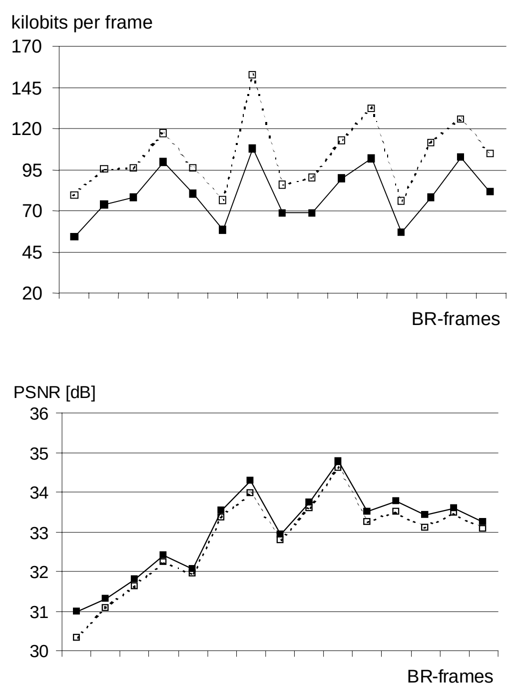


Figure 3. Bitrate (in kilobits per frame)and PSNR for improved and standard prediction of BR-frames: solid line – improved prediction,dotted line – standard prediction used in an MPEG-2 hierarchical coder.

Table1. The experimental results for GOPs with 75% B-frames.

		Funfair	*Flower Garden*	*Stefan*	*Cheer*	*Bus*	*Basketball*
Single layer coder (MPEG-2)	Bitsream [Mb]	5,18	5,27	5,14	5,21	5,19	5,17
	Average PSNR [dB] for luminance	32,18	30,97	35,09	31,93	34,57	31,36
Proposed scalable coder	Bitstream [Mb]	5,63	6,40	5,85	5,35	6,33	5,91
	Average PSNR [dB] for luminance	32,17	30,92	35,11	31,99	34,51	31,35
	Base layer bitstream [Mb]	2,16	2,17	2,14	2,15	2,15	2,14
	Base layer bitstream as percent of the total bitsream	38,4	33,9	36,6	40,2	34	36,2
	Scalability overhead [%]	8,7	21,5	13,8	2,7	22,0	14,3

6 CONCLUSIONS

It is shown that application of an appropriate combination of tools already available in MPEG-2 scalable profiles leads to improvement of spatially scalable coders which are mostly considered as inefficient. The approach described exploits the B-frame prediction schemes already proposed in [17] for the subband-based predictive coders.

References

[1] ISO/IEC International Standard 13818, "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information".

[2] B. Haskell, A. Puri, A. N. Netravali, "Digital Video: An Introduction to MPEG-2", New York: Chapman & Hall, 1997.

[3] ISO/IEC Draft International Standard 14496, "Information Technology - Generic Coding of Audiovisual Objects".

[4] K. Shen, E. Delp, "Wavelet based rate scalable video compression," IEEE Trans. Circuits Syst. Video Technology, vol. 9, pp. 109-122, February 1999.

[5] T. Tsunashima, J. Stampleman, V. Bove, "A scalable motion -compensated subband image coder," IEEE Trans. on Communication, vol. 42, pp. 1894-1901, Febr./March/April 1994.

[6] P.-Y. Cheng, J. Li, C.-C. Kuo, "Multiscale video compression using wavelet transform and motion compensation," in Proc. International Conference on Image Processing, vol. 1, pp. 606-609, 1995.

[7] U. Benzler, "Scalable multiresolution video coding using sub-band decomposition," in First Int. Workshop on Wireless Image/Video Communication, pp. 109-114, Loughborough, 1996.

[8] P.-C. Chang, T.-T. Lu, "A scalable video compression technique based on wavelet and MPEG coding," in Proc. Int. Conference on Consumer Electronics, pp. 372-373, 1999.

[9] F. Bosveld, "Hierarchical video compression using SBC," Ph.D. dissertation, Delft University of Technology, Delft 1996.

[10] Y.-Q. Zhang, S. Zafar, "Motion-compensated wavelet transform coding for color video compression," IEEE Trans. Circuits Syst. Video Technology, vol. 2, pp. 285-296, September 1992.

[11] H.Gharavi, W.Y.Ng, "H.263 Compatible Video Coding and Transmission," in Proc. First International Workshop on Wireless Image/Video Communication, pp. 115-120, Loughborough 1996.

[12] U. Benzler, "Scalable multi-resolution video coding using a combined subband-DCT approach," in Proc. Picture Coding Symposium, pp. 17-20, Portland 1999.

[13] M. Domański, A. Łuczak, S. Maćkowiak, R. Świerczyński, "Hybrid coding of video with spatio-temporal scalability using subband decomposition," in Signal Processing IX: Theories and Applications, pp. 53-56, Rhodes, September1998.

[14] M. Domański, A. Łuczak, S. Maćkowiak, R. Świerczyński, "Hybrid coding of video with spatio-temporal scalability using subband decomposition," in Proceedings of SPIE, vol. 3653, pp. 1018-1025, San Jose CA, 1999.

[15] M. Domański, A. Łuczak, S. Maćkowiak, R. Świerczyński, U. Benzler, "Spatio-temporal scalable video codecs with MPEG-compatible base layer," in Proc. Picture Coding Symposium, pp. 45-48, Portland 1999.

[16] G. Conklin, S. Hemami, ""A comparison of temporal scalability techniques," IEEE Trans. Circuits Syst. Video Technology, vol. 9, pp. 909-919, September 1999.

[17] M. Domański, A. Łuczak, S. Maćkowiak, "Scalable MPEG video coding with improved B-frame prediction", Proceedings IEEE Int. Symp. Circuits and Systems, Geneva 2000, in print.