

# SCALABLE MPEG VIDEO CODING WITH IMPROVED B-FRAME PREDICTION

Marek Domański, Adam Łuczak, Sławomir Maćkowiak

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

## ABSTRACT

Recently, there is a great interest in video codecs that implement the functionality of spatial scalability. Unfortunately, those MPEG-2 and MPEG-4 coders which exhibit such a functionality produce much more bits than corresponding single layer coders. This bitrate overhead can be reduced by application of spatio-temporal scalability as proposed by the authors. The base layer bitstream corresponds to pictures with reduced both spatial and temporal resolution while the enhancement layer bitstream is used to transmit the information needed to retrieve images with full spatial and temporal resolution. Full compatibility with the MPEG standards is ensured in the base layer where temporal resolution reduction is obtained by B-frame data partitioning, i.e. by placing each second frame (B-frame) in the enhancement layer only. Improved prediction of B-frames in the enhancement layer is proposed in this paper. The idea is to combine temporal forward and backward prediction with spatial interpolation. Experimental results prove clear improvement of the MPEG-2-compatible scalable coding efficiency for the scheme proposed.

## 1. INTRODUCTION

Development of communication networks with different transmission bitrates as well as a demand for better protection of video transmission in error-prone environments has stimulated research on spatially scalable or hierarchical video coders. Such coders produce two bitstreams: a base layer bitstream which represents low resolution pictures and an enhancement layer bitstream which provides additional data needed for reproduction of pictures with full resolution. An important feature is that the base layer bitstream can be decoded independently from the enhancement layer.

The functionality of spatial scalability 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.

There were many attempts to improve spatially scalable coding of video. Among various proposals, application of subband decomposition should be considered as very promising [4-9]. The idea is to split each image into four spatial subbands. The subband LL of lowest frequencies

constitutes a base layer while the other three subbands are jointly transmitted in an enhancement layer (Fig. 1). Unfortunately, in most of such coders, it is difficult to allocate appropriate number of bits to the base layer and to the enhancement layer. 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, spatio-temporal scalability has been proposed by the authors [10,11]. Here, a base layer corresponds to the bitstream of the pictures with reduced both spatial and temporal resolutions. Therefore, in the base layer, the bitrate is decreased as compared to encoding with spatial scalability only. Now, it is easy to get the base layer bitrate equal or even less than that of the enhancement layer. The enhancement layer is used to transmit the information needed for restoration of the full spatial and temporal resolution.

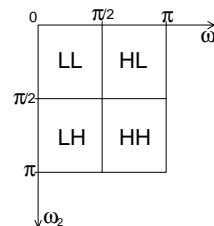


Figure 1. Subband decomposition

The goal of our work is to propose alternative techniques, which would provide spatial scalability with lower bitrate overheads. The assumption is that high level of compatibility with the MPEG video coding standards would be ensured. In particular, it is assumed that the low-resolution base layer bitstream is fully compatible with the MPEG standards.

We are going to prove that these goals can be achieved using spatio-temporal scalability with improved prediction of B-frames. For the sake of simplicity the technique will be described in the terms of the MPEG-2 standard only but it is applicable also to MPEG-4 encoding of video.

## 2. SPATIO-TEMPORAL SCALABILITY WITH B-FRAME DATA PARTITIONING

We assume that the number of B-frames between two consecutive I- or P-frames is odd. Temporal resolution

reduction is achieved by partitioning of the stream of B-frames: each second frame is included into the enhancement layer only. Therefore we have 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). The base layer represents the subband LL from I-, P- and BR-frames, and the enhancement layer represents BE-frames, subbands LH, HL, HH from I-, P-frames and hierarchical enhancement of the BR-frames.

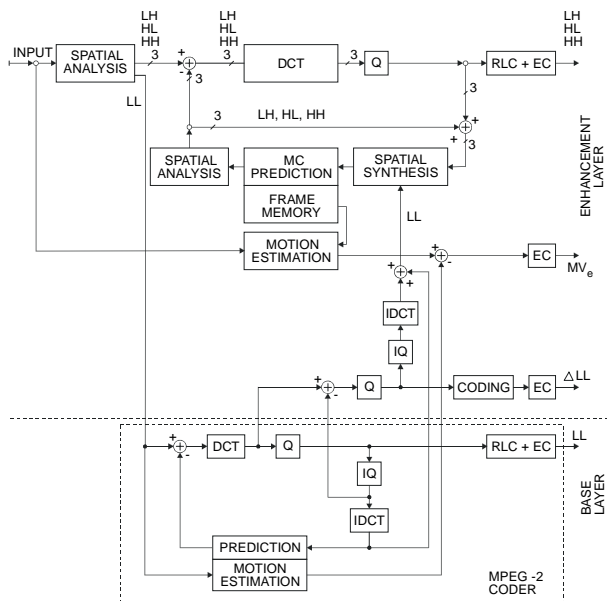


Figure 2. Block diagram of the coder.

Base layer coder is implemented as a motion-compensated hybrid MPEG-2 coder. In the enhancement layer coder, motion is estimated for full-resolution images and full-frame motion compensation is performed. Therefore all subbands have to be synthesized into full frames. After motion compensation spatial subbands are produced again. The prediction errors are calculated and encoded for three subbands (HL, LH, and HH) of I- and P-frames.

Table 1. Frames of different types in both layers.

Base	I	BR	P	BR	P	BR	P		
Enhancement	I	BE	BR	BE	P	BE	BR	BE	P

Motion vectors  $MV$  are transmitted for the base layer. Another motion vectors are estimated for the enhancement layer. In the enhancement layer, difference values  $MV_e$  are transmitted. The overall structure of the coder for P-frames is shown in Fig. 2. Intraframe coding is performed using a part of this system. Unfortunately, encoding of B-frames in the interframe coder from Fig. 2 is not efficient because the bitstream  $\Delta LL$  constitutes too large portion of the whole bitstream related to B-frame.

In the enhancement layer coder, the subband LL used for frame synthesis is more finely quantized than this transmitted in the base layer. It corresponds to a sum of information contained in the base layer and in the bitstream  $\Delta LL$  transmitted in the enhancement layer. The bitstream  $\Delta LL$  contains bitplanes correcting the transform coefficients transmitted in the base layer.

### 3. IMPROVED B-FRAME PREDICTION

Improved prediction is proposed for the BR-frames, which are the B-frames represented in both layers. Each macroblock in a full-resolution BR-frame can be predicted from the following reference frames (Fig.3):

- previous reference frame RP (I- or P-frame),
- next reference frame RN (I- or P-frame),
- current reference frame RC (BR-frame).

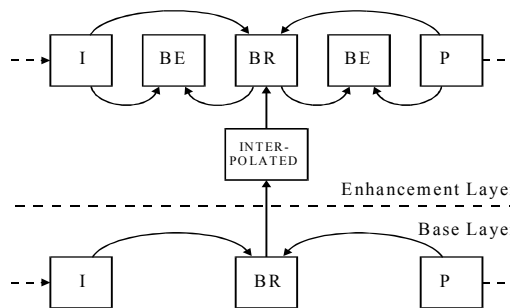
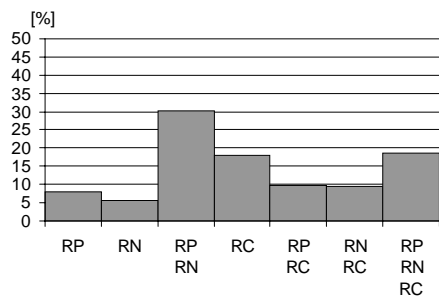
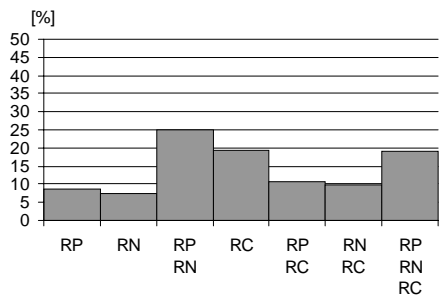
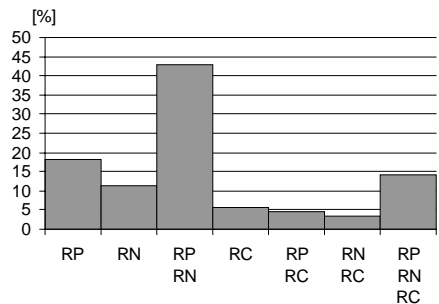
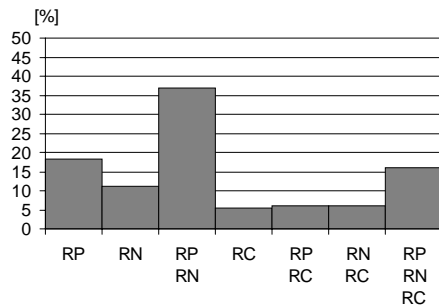


Figure 3. Improved prediction of B-frames.

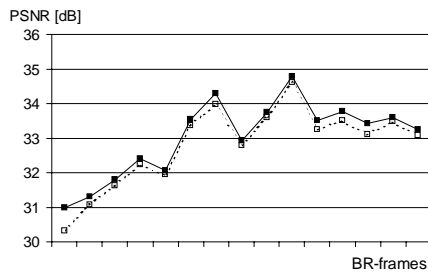
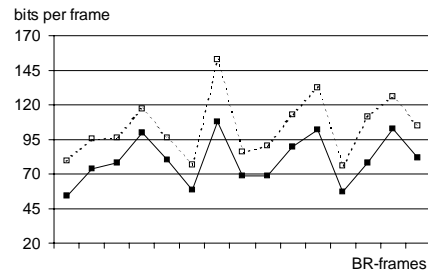
The data from the previous and next reference frames RP and RN are motion-compensated, and data from the current reference frame are upsampled in the two-dimensional space domain. The best suited reference frame or average of two or three reference frames is chosen according to the criterion of smallest prediction error.

The improvement on standard MPEG prediction within a single layer consists in usage of three reference frames instead of two reference frames allowed in the MPEG standards. Experimental results with television test sequences prove that a current reference frame, i.e. a low-resolution BR-frame, is used in prediction of significant portion of macroblocks, sometimes even for more than 50% of all macroblocks in BR-frames (Fig. 4). Application of the improved prediction leads to lower bitrates and higher PSNR as compared to standard prediction (Fig. 5).

The explanation of the results is that improved prediction replaces I-macroblocks with macroblocks predicted using current reference frame RC (Fig. 6). Of course, much less bits are needed to encode the latter macroblocks than macroblocks of type I.



**Figure 4.** Percentage of macroblocks predicted using individual reference frames or their averages. From the top to the bottom: the test sequence *Flower garden* (15 frames) encoded for 5,24 and 8,76 Mbps and the test sequence *Funfair* (15 frames) encoded for 5,03 and 8,48 Mbps.



**Figure 5.** Bitrate and PSNR for improved and standard prediction of BR-frames: solid line – improved prediction, dotted line – standard prediction.

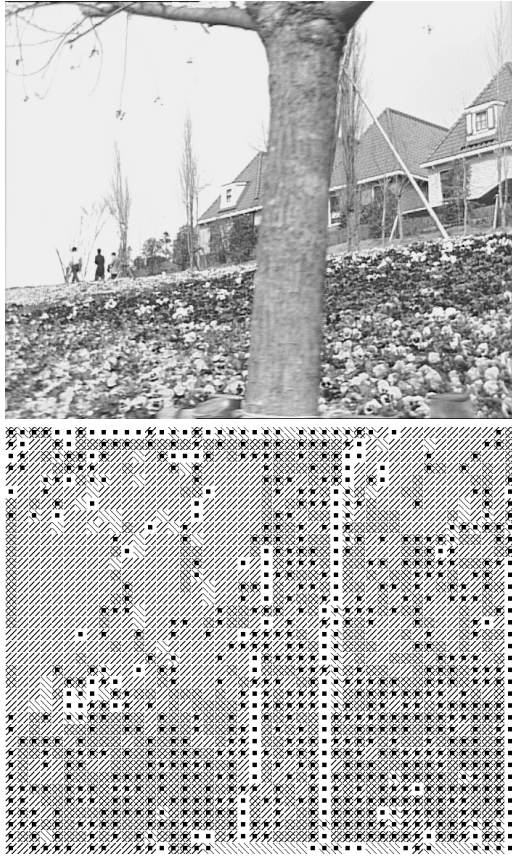
The difference from the MPEG-2 spatially scalable profile consists in choosing the best prediction using all three reference frames simultaneously.

Moreover, BR-frames are used as reference frames for BE-frames in the enhancement layer. Therefore prediction of BE-frames is more precise (Fig. 3).

#### 4. CONCLUSIONS

The experimental results for progressive  $720 \times 576$ , 50 Hz test sequences „Flower garden” and „Funfair” are given in Table 3.

The experimental results from Table 2 prove high efficiency of the coder. With the same bitrate as by MPEG-2 non-scalable profile, the scalable coder proposed reaches almost the same quality. The codec proposed significantly outperforms spatially scalable MPEG-2 [1] coders which generate bitrate overheads mostly exceeding 50%.



**Figure 6.** A frame from the test sequence *Flower garden* and various types of macroblocks in the corresponding BR-frame: slash signs denote forward prediction, backslash signs denote backward prediction and square signs denote interpolation.

## 5. 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] International Organization For Standardization, "MPEG-4 Video Verification Model", ISO/IEC JTC1/SC29/WG11, Stockholm, 1998.
- [4] T. Tsunashima, J. Stampleman, V. Bove, "A scalable motion -compensated subband image coder," *IEEE Trans. on Communication*, vol. 42, pp. 1894-1901, 1994.
- [5] F. Bosveld, "Hierarchical video compression using SBC," Ph.D. dissertation, Delft University of Technology, Delft 1996.
- [6] 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.
- [7] Senbel, H. Abdel-Wahab, "Scalable and robust image compression using quadtrees," *Signal Processing: Image Communication*, vol. 14, pp. 425-441, 1999.
- [8] Shen, E. Delp, "Wavelet based rate scalable video compression," *IEEE Trans. Circuits Syst. Video Technology*, vol. 9, pp. 109-122, February 1999.
- [9] U. Benzler, "Scalable multi-resolution video coding using a combined subband-DCT approach", *Picture Coding Symposium 1999*, pp. 17-20.
- [10] 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, September 1998.
- [11] M. Domański, A. Łuczak, S. Maćkowiak, R. Świerczyński, Ulrich Benzler „Spatio-Temporal Scalable Video Codecs with MPEG-Compatible Base Layer”, *Picture Coding Symposium 1999*, pp. 45-48.

**Table 2.** Experimental results.

		Test sequence			
		<i>Flower Garden</i>		<i>Funfair</i>	
Single Layer (MPEG-2)	Bitsream [Mb]	5.24	8.76	5.03	8.48
	Average luminance PSNR [dB]	30.63	33.75	31.38	34.07
Proposed scalable coder	Bitsream [Mb]	5.34	8.74	5.04	8.48
	Average luminance PSNR [dB]	30.44	33.41	30.81	34.16
	Base layer bitstream [Mb]	2.21	3.00	2.11	2.70
	Base layer bitstream [%]	41	34	41	31