SPATIO-TEMPORAL SCALABILITY FOR MPEG-2 VIDEO CODING

Adam Łuczak, Sławomir Maćkowiak, Marek Domański

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

EXTENDED SUMMARY

1. STATEMENT OF THE PROBLEM

Spatially scalable or hierarchical video 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. The base layer bitstream can be decoded independently from an enhancement layer. The functionality of spatial scalability is very important for video transmission through inhomogeneous communication networks, i.e. networks with different transmission bitrates as well as for better protection of video transmission in error-prone environments.

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. Unfortunately, in most of such coders, it is difficult to allocate appropriate number of bits to the base layer and to the enhancement layer.

2. SPATIO-TEMPORAL SCALABILITY

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.

3. GOALS

The goal of the work is to achieve total bitrate of both layers of scalable coding 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 particular, it is assumed that the low-resolution base layer bitstream is fully compatible with the MPEG-2 standard.

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.

4. SPATIO-TEMPORAL SCALABILITY WITH B-FRAME DATA PARTITIONING

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. 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.

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.

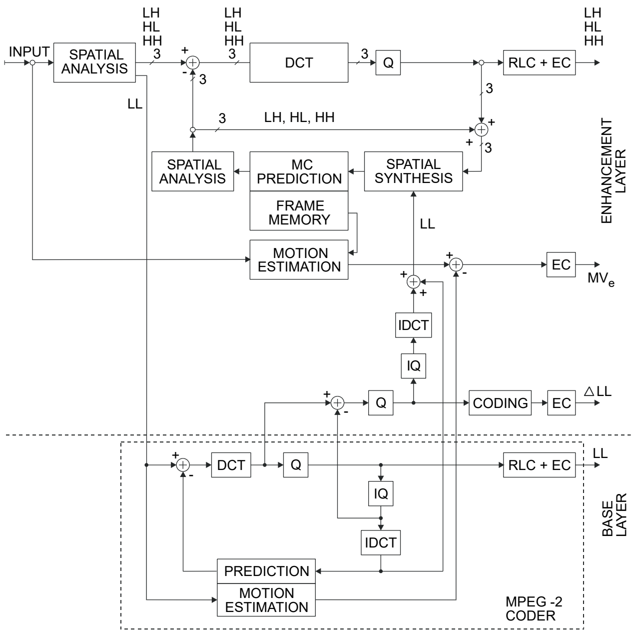

Figure 2. Block diagram of the coder.

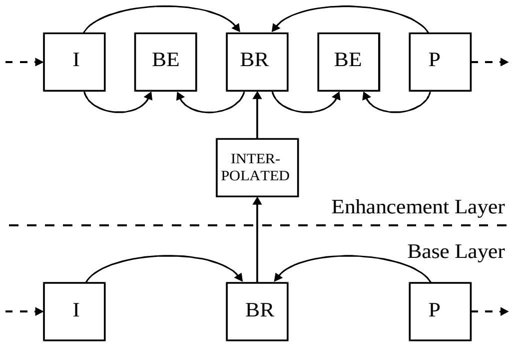

Figure 3. Improved prediction of B-frames.

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 ΔLL 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 ΔLL transmitted in the enhancement layer. The bitstream ΔLL contains bitplanes correcting the transform coefficients transmitted in the base layer.

5. 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).

The improvement on standard MPEG spatially scalable coding consists in usage of three reference frames (Fig. 4) 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.

6. EXPERIMENTAL RESULTS

The purpose of the experiments was to estimate the best coder structure and its properties. Therefore software was written in C++ language. The most important feature is its flexibility allowing tests of different variants of coding algorithm. Currently the program includes about 13 000 lines of code. It includes software implementation of an MPEG-2 MP@ML coder for the base layer. The software runs on Sun 20 workstations under the Solaris operational system. The coder is aimed at processing of progressive 720×576, 50 Hz, 4:2:0 test sequences.

Table 1.Experimental results (*abbreviated*).

		Test sequence			
		Flower Garden		*Funfair*	
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	Bitstream [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

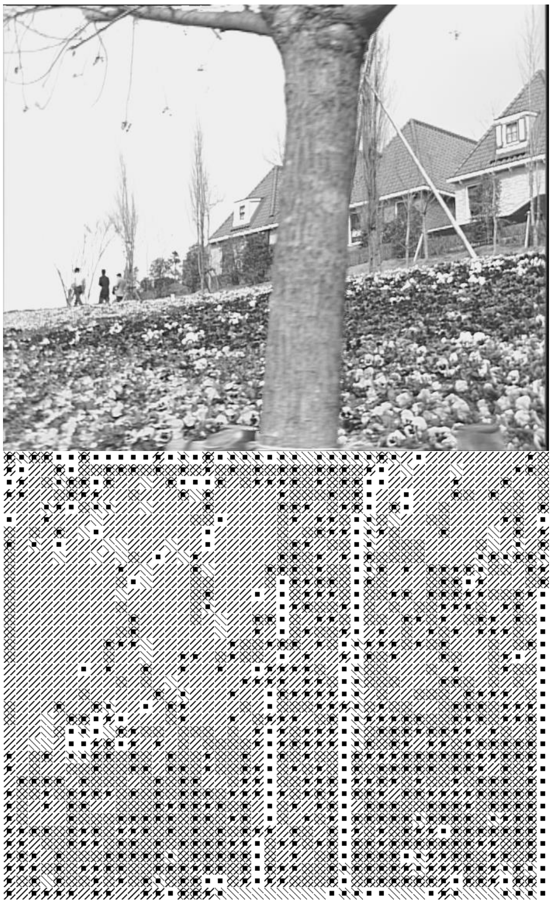

Figure 4. A frame from the test sequence *Flower garden* and various types of macroblocks in the corresponding BR-frame: slash signs denote forward prediction, backslash signs denote backward prediction and square signs denote interpolation.

7. CONCLUSIONS

The experimental results from Table 2 prove high efficiency of the coder. With the same bitrate as by MPEG-2 nonscalable 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%.

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.