Digital Signal Processing for Multimedia Communications and Services, EURASIP ECMCS 2001, Budapest, September 11-13 2001, pp. 189-192.

Three-Layer MPEG-2 — Based Video Codecs with Spatio-Temporal Scalability

M. Domański and S. Maćkowiak*
Institute of Electronics and Telecommunications
Poznan University of Technology
Piotrowo 3A, Poznań, Poland
Phone: (48) 61-665 2762 Fax: (48) 61-665 2572 E-mail: domanski@et.put.poznan.pl

Keywords: Video compression, scalable video coder, MPEG 2, fine granularity.

Abstract - The paper deals with a three-layer video coder based on spatio-temporal scalability and data partitioning. The proposed coder consists of two parts: a low-resolution coder and full-resolution coder. The first part encodes pictures with reduced spatial and temporal resolution. The full-resolution encoder exploits interpolated images from the base layer. The enhancement layer data are partitioned into two bitstreams with fine granularity. The base layer encoder is fully compatible with the MPEG-2 video coding standard.

I. INTRODUCTION

There is a growing demand for multilayer scalable video codecs that are suitable for video transmission over heterogeneous communication networks characterized by various available levels of quality of service. Different levels of quality of service are mostly related to different available transmission bitrates. For this purpose, the transmitted bitstream has to be partitioned into some layers in such a way that an arbitrary number of layers is decodable into a video sequence with reduced spatial resolution, temporal resolution or signal-to-noise ratio (SNR). This functionality is called scalability [1,2]. An important feature of scalable bitstreams is that a base layer bitstream can be decoded independently from other layers. Therefore low-resolution terminals are able to decode only the base layer bitstream in order to display low-resolution pictures. Terminals connected to network characterized by full quality of service are able to decode all layers.

Scalable coding would not be acceptable for generic applications, if the quality is significantly worse than the quality achievable with single-layer coding. It would never be accepted if it is in any way less efficient than simulcast coding. Application of spatial scalability alone is not practical because halving both (horizontal and vertical) spatial resolutions results usually in bitrate reduction of about 30% only. Among various proposals for spatially scalable coding of video, application of subband/wavelet decomposition should be considered as very promising [4,6,7,8]. Unfortunately, in most of such coders, it is difficult to allocate appropriate number of bits to the layers. Universal scalable coding would require flexible combinations of spatial/temporal/SNR scalability with fine granularity [10]. Therefore a combination of spatial and temporal scalabilities has been proposed by the authors [11,12,13]. Such a two-layer system exhibits moderate bitrate overheads due to scalability. Straightforward extension of this idea onto three-layer systems is possible but not very practical because with an input of resolution 720×576 pixels and 50 frames per second, the resolution of the base layer would be QCIF (180×144) with 12.5 frames per second. Therefore a combination of spatio-temporal scalability with DCT data partitioning related also to some SNR scalability is proposed in this paper.

The concept of the codec is derived from the idea of two-layer coding. This presents a proposed modification of the two-layer system based on mixed spatial and temporal scalability by additional data partitioning in the enhancement layer [15]. The assumption for the proposal is that high level of compatibility with the MPEG video coding standards would be ensured. In the paper, the MPEG-2 [1,2] video coding standard is used as reference but the results are also applicable to the MPEG-4 [3] systems with minor modifications.

The goal of the experiments was to verify that the proposed data partitioning between the middle and the enhancement layers is efficient.

II. THREE-LAYER CODER STRUCTURE

The proposed three-layer coder consists of two parts:

- The low resolution coder which produces base layer bitstream with reduced spatial and temporal resolutions. Temporal resolution reduction is achieved by partitioning of the stream of B-frames: each second frame is not included into the base layer. Separable FIR filter banks are used for spatial decimation and interpolation.
- The full resolution coder which produces a bitstream which is partitioned between the middle and the enhancement layers. The middle layer represents video with full spatial and temporal resolution but reduced quality (reduced PSNR) obtained by DCT data partitioning, i.e. leaving some high-frequency DCT coefficients for the enhancement layer. The enhancement layer consists the rest of DCT coefficients necessary to restore full quality video with full temporal and spatial resolutions.

The middle layer comprises all headers, motion vectors and low-frequency DCT coefficients while the enhancement layer comprises the rest of DCT coefficients as well as slice headers that are needed for resynchronisation in the case of uncorrected transmission errors.

* Mr. S. Maćkowiak is Fellow of the Foundation for Polish Science.

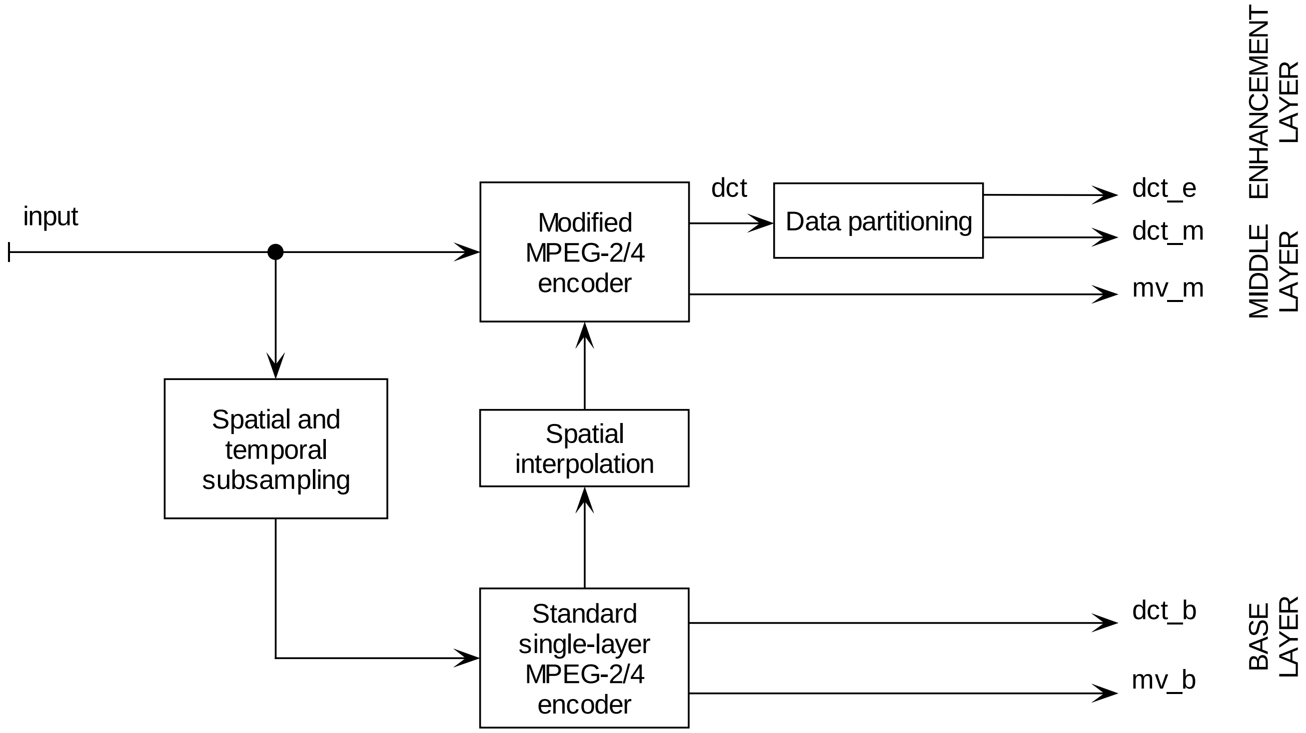

Fig. 1. The general structure of a three-layer coder
(*dct_b* and *mv_b* – base layer, *dct_m* and *mv_m* – middle layer, *dct_e* – enhancement layer).

The ratio of bitrates between middle and enhancement layers is directly controlled by the number of nonzero DCT coefficients allocated to the middle layer. The amount of DCT data in the middle layer is almost proportional to the number of nonzero coefficients allocated to this layer.

The bitstream produced in the base layer is described by fully compliant MPEG-2 standard syntax. The motion vectors *mv_b* for the low-resolution images are estimated independently from those vectors estimated for the enhancement layer. Motion vectors *mv_b* are transmitted for the base layer. The other part of the coder produces bitstreams for both middle and enhancement layer. In particular, motion is estimated for full-resolution images and full-frame motion compensation is performed. Therefore the number of motion vectors *mv_m* sent in the enhancement layer is four times that of the base layer.

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

III. EXPERIMENTAL RESULTS

The verification model of the three-layer scalable coder has been prepared as software written in C++ language. The most important feature is its flexibility allowing tests of different variants of coding algorithm. The software runs on PC compatible computers under the Windows NT operational system. The coder is implemented for processing of progressive 4:2:0 720 × 576 sequences with 50 frames per second. The base layer coder is a standard MPEG 2 coder that processes video in the SIF format but both the middle layer and the enhancement layer are in the full Standard Definition TV (SDTV) resolution.

The goal of the experiments was to verify that the proposed data partitioning between the middle and the enhancement layers is effective.

Simulations are carried out for constant quality coding, corresponding to approximately 5 and 7 Mbps for non–scalable MPEG–2 coding of SDTV signals. Table 1 contains an experimental results for the three-layer scalable coder. In this case middle layer comprises all headers, motion vectors and one DCT coefficient per block. The enhancement layer consists of the rest of DCT coefficients necessary to restore full quality video. Such data partitioning causes that the coder produces three bitstreams with similar bitrates. Experimental results prove that each bitstream constitutes about 30÷40% of the overall bitstream.

Figures 2, 3, 4 and 5 show number of bits and PSNR ratings for particular frame in two GOPs encoded test sequence. The results for test sequence “FunFair” are presented on figures 2 and 3. Compared with non-scalable MPEG-2 coding at bitrate 5.2Mbps, the increase is between 9% and 16%. The results for test sequence “Cheer” are presented on figures 4 and 5. Compared with non-scalable MPEG-2 coding at bitrate 7Mbps, the increase is between 3% and 12%. The bitrate overhead between 3% and 16 % measured relative to the single layer MPEG-2 bitstream is acceptable for generic applications.

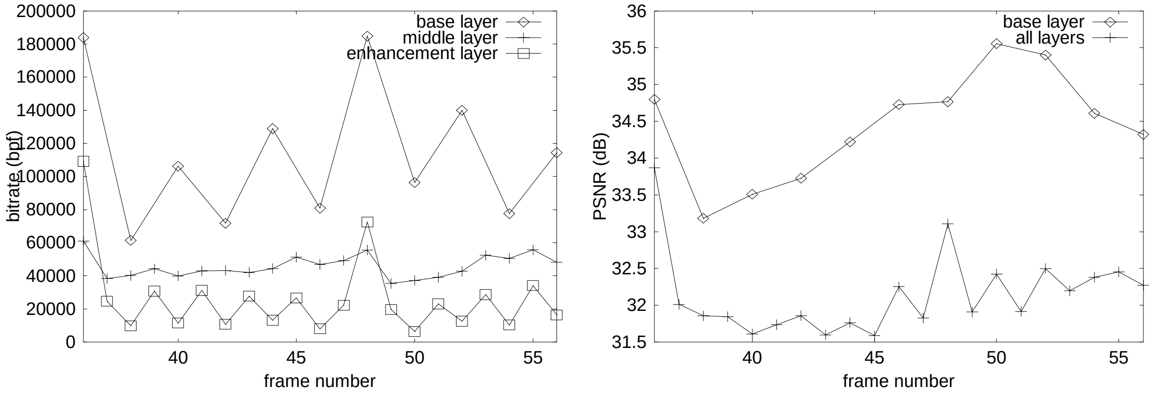

Fig. 2. Bits per frame and PSNR for two GOPs of the test sequence FunFair 5.7Mbps.

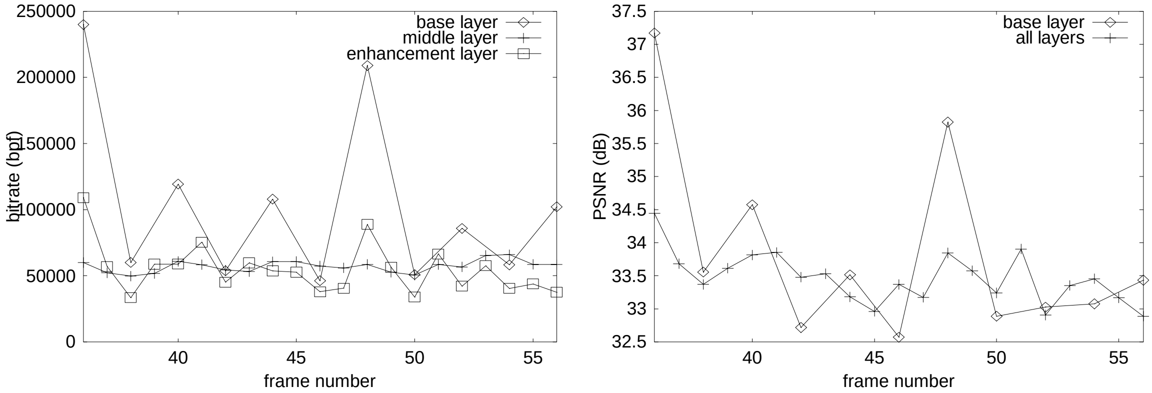

Fig. 3. Bits per frame and PSNR for two GOPs of the test sequence FunFair 8.1Mbps.

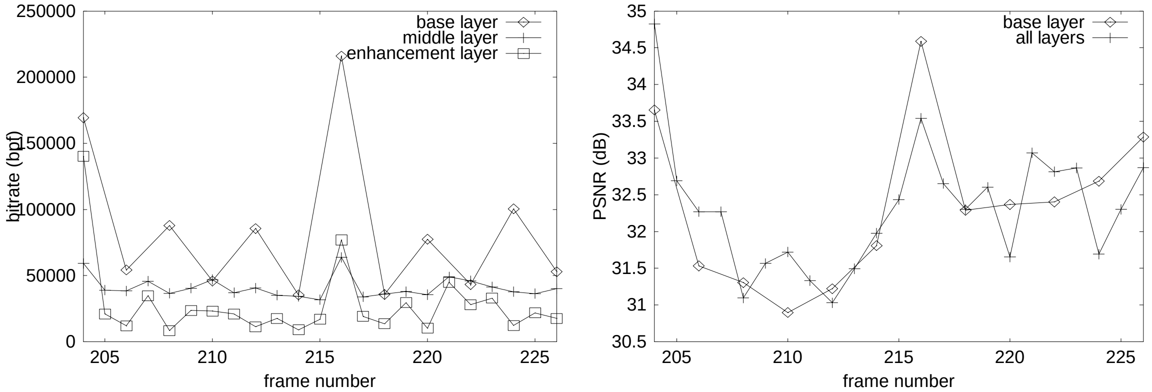

Fig. 4. Bits per frame and PSNR for two GOPs of the test sequence Cheer 5.4Mbps.

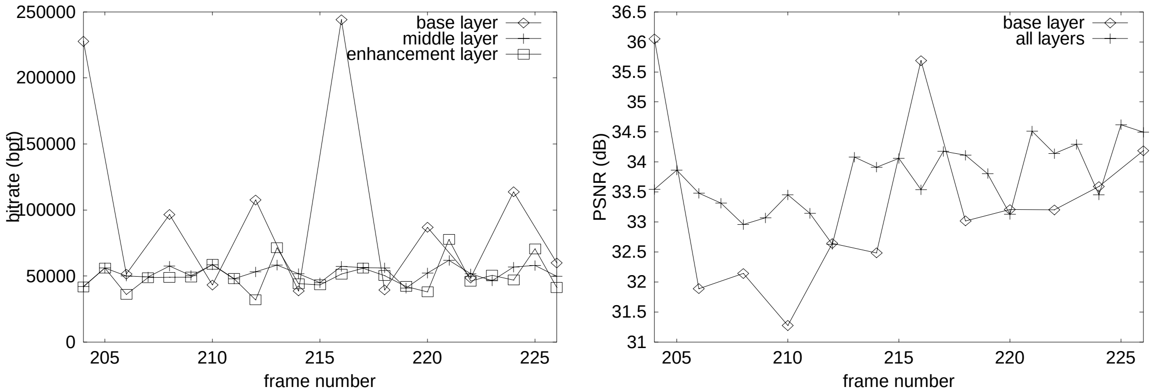

Fig. 5. Bits per frame and PSNR for two GOPs of the test sequence Cheer 7.8Mbps.

TABLE I
EXPERIMENTAL RESULTS FOR BT.601 PROGRESSIVE SEQUENCES FUNFAIR AND CHEER

		Funfair		*Cheer*	
Single layer coder (MPEG-2)	Bitstream [Mb]	5.2	7.0	5.2	7.0
	Average PSNR [dB] for luminance	32.2	33.8	31.9	33.7
Proposed three-layer coder	Base layer average PSNR [dB] for luminance	33.1	34.1	31.7	33.1
	Average PSNR [dB] for luminance recovered from all three layers	32.2	33.8	32.0	33.7
	Base layer bitstream [Mb]	2.2	2.5	2.1	2.5
	Middle layer bitstream [Mb]	2.2	2.8	2.0	2.7
	Enhancement layer bitstream [Mb]	1.3	2.8	1.3	2.6
	Total bitstream [Mb]	5.7	8.1	5.4	7.8
	Scalability overhead [%]	9.6	15.7	3.8	11.1

IV. CONCLUSION

In this article, the three-layer video codec that provides spatio-temporal scalability mixed with data partitioning has been proposed. The experimental data prove that the coder is able to produce three bitstreams with similar bitrates. Such bit allocation is very advantageous for practical applications. The experimental results prove that contrary to one scalability mode alone acceptable overhead can be achieved by flexible combinations of spatial/temporal/SNR scalability with fine granularity.

REFERENCES

[1] ISO/IEC IS 13818, Generic Coding of Moving Pictures and Associated Audio Information.

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

[3] ISO/IEC IS 14496, Generic Coding of Audiovisual Objects.

[4] T. Tsunashima, J. Stampleman, V. Bove, “A Scalable Motion-Compensated Subband Image Coder”, *IEEE Trans. Communication*, 42 (1994) 1894-1901.

[5] H. Gharavi, W.Y. Ng, “H.263 Compatible Video Coding and Transmission”, *Proc. First International Workshop on Wireless Image/Video Communication*, Loughborough (1996) 115-120.

[6] F. Bosveld, *Hierarchical Video Compression Using SBC*, Ph.D. dissertation, Delft Univ. of Technology, Delft (1996).

[7] S. Senbel, H. Abdel-Wahab, “Scalable and Robust Image Compression Using Quadtrees”, *Signal Processing: Image Communication* 14 (1999) 425-441.

[8] K. Shen, E. Delp, “Wavelet Based Rate Scalable Video Compression”, *IEEE Trans. Circ. Syst. Video Technology* 9 (1999) 109-122.

[9] P.-C. Chang, T.-T. Lu, “A Scalable Video Compression Technique Based on Wavelet and MPEG Coding”, *Proc. Int. Conf. on Consumer Electronics* (1999) 372-373

[10] U. Benzler, “Spatial Scalable Video Coding Using a Combined Subband-DCT Approach”, *IEEE Trans. Circ. Syst. Video Technology* 10 (2000) 1080-1087.

[11] M. Domański, A. Łuczak, S. Maćkowiak, R. Świerczyński, “Hybrid Coding of Video with Spatio-Temporal Scalability Using Subband Decomposition”, *Signal Processing IX: Theories and Applications*, Rhodes (1998) 53-56.

[12] M. Domański, A. Łuczak, S. Maćkowiak, “Spatio-Temporal Scalability for MPEG Video Coding”, *IEEE Trans. Circ. Syst. Video Technology* 10 (2000) 1088-1093.

[13] A. Łuczak, S. Maćkowiak, M. Domański, “Spatio-Temporal Scalability Using Modified MPEG-2 Predictive Video Coding”, *Signal Processing X: Theories and Applications*, Tampere (2000) 961-964.

[14] M. Domański, A. Łuczak, S. Maćkowiak, “On Improving MPEG Spatial Scalability”, *Proc. Int. Conf. Image Proc.*, Vancouver (2000) 848-851.

[15] M. Domański, A. Łuczak, S. Maćkowiak, “Modified MPEG-2 video coders with efficient multi-layer scalability”, *Proc. IEEE Int. Conf. Image Proc.*, Tessaloniki, 2001, submitted for publication.