AN MPEG – BASED MULTILAYER VIDEO CODER WITH SPATIO – TEMPORAL SCALABILITY AND DCT DATA PARTITIONING

Marek Domański, Sławomir Maćkowiak*

Institute of Electronics and Telecommunication, Poznań University of Technology
ul. Piotrowo 3a, 60-965 Poznań, Poland, email: [domanski][smack]@et.put.poznan.pl

Abstract: The paper deals with a new three-layer video coder based on spatio-temporal scalability and DCT data partitioning. The base layer is fully MPEG-2 (or MPEG-4) compatible and the whole structure exhibits high level of compatibility with individual building blocks of MPEG-2/4 coders. The bitrate overhead measured relative to the single layer MPEG-2 bitstream varies between 5% and 25% for progressive television test sequences. The paper deals with the problem of control of bitrate allocation between layers for the constant bitrate mode of operation. The experimental data prove that bitrates can be flexibly controlled in order to match communication channel requirements.

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

1. INTRODUCTION

Recent developments in video compression standards are mostly aimed at new functionalities rather than at increase of coding efficiency alone. This priority change has became clearly remarkable during the standardization procedures of MPEG-4 [1] and JPEG2000 [2]. Among many functionalities provided by new international standards, scalability is that of paramount importance. The bitstream produced by scalable video coders is partitioned into at least two parts: an independently decodable base layer bitstream that represents low-quality video and an enhancement layer bitstream which provides additional data needed for reproduction of full-quality video. Scalability is substantial for video communication via heterogeneous networks, video transmission in error-prone environments as well as broadcasting to terminals defined in various standards.

The base layer represents a video sequence with reduced spatial resolution, temporal resolution or signal-to-noise ratio (SNR). The respective functionality is called spatial, temporal or SNR scalability. Another similar mechanism called data partitioning consists in allocation of the crucial data to the base layer while the other data belong to the enhancement layer.

All the above mentioned tools are already standardized in MPEG-2 [3,4] and MPEG-4 [1]. Unfortunately, the standard solutions for spatial scalability suffer from low efficiency related to high bitrate overheads as compared to single-layer coding. On the other hand,

---

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

spatial scalability called also hierarchical compression has been implemented in a very efficient way using wavelet decomposition of still images. The approaches to scalable video compression recently reported in the references are mostly related to SNR scalability or wavelet decomposition [5-9].

2. PROBLEM CONSIDERED

In contrary to most of the approaches described in the references, this paper deals with a proposal where a two-layer system [10] based on mixed spatial and temporal scalability is combined with data partitioning resulting in a three-layer system [11,12]. 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. The work is aimed at video coding for bitrates of few Mb/s.

Application of spatial scalability alone is not practical because halving both spatial resolutions (horizontal and vertical) results usually in bitrate reduction about 30% only. Therefore a combination of spatial and temporal scalabilities has been proposed by the authors [10]. 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 an input resolution of 720×576 pixels and 50 frames per second would imply the base layer in QCIF format (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 problem considered in this paper is related to bitrate control in the individual layers. For the practical reasons, in the constant bitrate mode of operation, the bitrates of the individual layers must be controllable in some range. The range of efficient bitrate control should be around the point where all three bitrates are equal.

3. CODER STRUCTURE

The proposed three-layer coder produces three bitstreams:

- Base layer with reduced spatial and temporal resolutions. The numbers of rows and columns of pixels in a frame as well as the frame frequency are halved.
- Middle layer 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.
- Enhancement layer providing the rest of DCT coefficients necessary to restore full quality video with full temporal and spatial resolutions.

The coder (cf. Fig. 1) consists of two motion-compensated video coders:

- Base-layer encoder is a standard single-layer MPEG-2 (or MPEG-4) encoder that performs compression of video with reduced spatial and temporal resolution.
- Middle and enhancement-layer encoder that performs compression of full-resolution video. This encoder performs independent motion estimation and exploits interpolated reconstructed base-layer images in order to avoid simulcast coding.

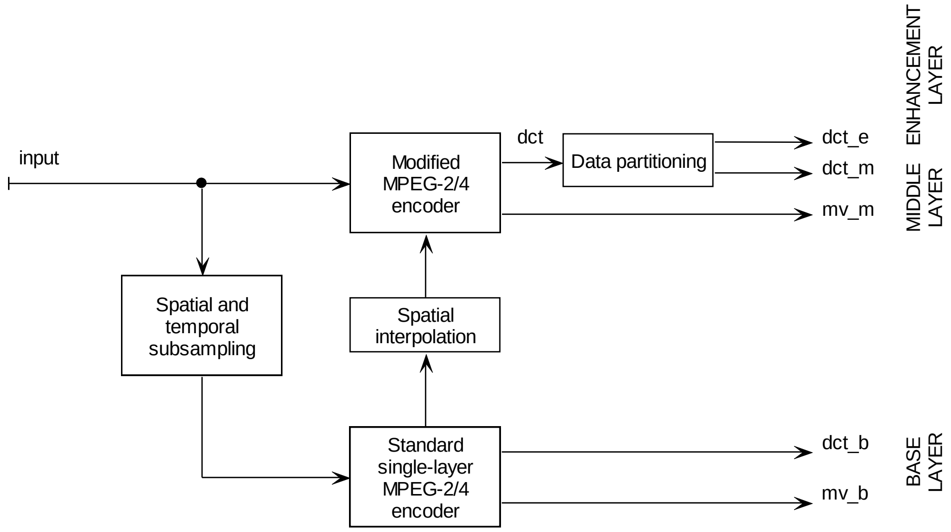

Fig. 1. Block diagram of the three-layer coder.

The second encoder produces a bitstream that is partitioned between the middle layer and the enhancement layer. 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 slice headers that are needed for resynchronisation in the case of uncorrected transmission errors.

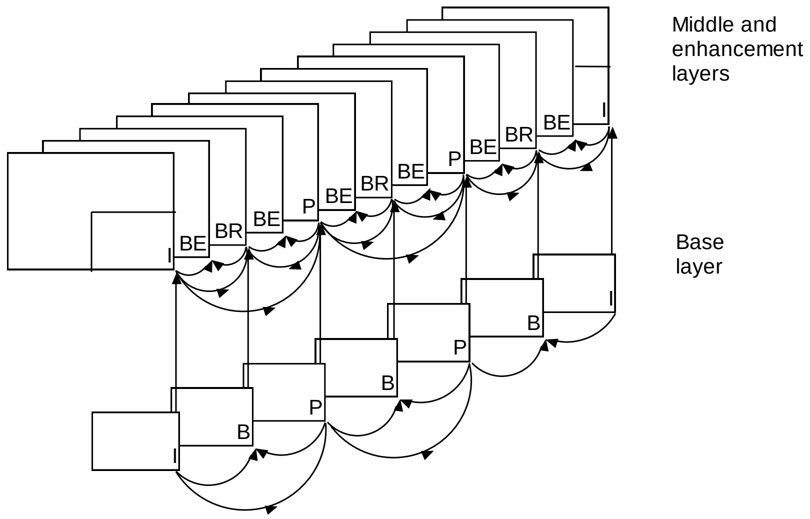

Fig. 2. The structure of a group of pictures. Horizontal arrows represent temporal prediction while vertical ones correspond to spatial interpolation.

Separable FIR filter banks are used for spatial decimation and interpolation. Temporal resolution reduction is obtained via B-frame data partitioning. Each second frame called BE-frame is skipped from the base layer thus existing in the enhancement layer only. The B-frames that exist both in the base and enhancement layers are called BR-frames. These frames are used as reference frames for temporal prediction of BE-frames in the enhancement layer (Fig. 2).

4. EXPERIMENTAL RESULTS

The verification model 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 aimed at processing of progressive 720×576, 50 Hz test sequences. Therefore the base layer is in the SIF format but both the middle layer and the enhancement layer are in the full television resolution. The experiments have been performed with 4:2:0 sequences as the verification model implements the coder that is an augmented version of main profile / main level MPEG 2 coder.

For the sake of brevity, the results are given for one video sequence only but similar results have been obtained for other test video sequences. Table 1 proves that similar bitrates can be achieved in the three layers for some reasonable total bitrate range. The results are given for 50Hz progressive video that corresponds to somewhat higher bitrates than common 25Hz interlaced video.

Table 1. The experimental results for BT.601 progressive sequence *Fun*.

Single layer coder (MPEG-2)	Bitstream [Mb]	4.0	5.18	6.0	7.0
	Average PSNR [dB] for luminance	30.6	32.2	33.0	33.8
Proposed scalable coder	Base layer average PSNR [dB] for luminance	30.5	33.0	34.1	34.1
	Average PSNR [dB] for luminance recovered from two layers	28.5	29.7	29.6	29.7
	Average PSNR [dB] for luminance recovered from all three layers	30.6	32.1	33.0	33.8
	Base layer bitstream as percent of the total bitstream	31.8	37.6	37.0	31.0
	Middle layer bitstream as percent of the total bitstream	45.5	38.0	35.8	34.7
	Enhancement layer bitstream as percent of the total bitstream	22.7	24.4	27.2	34.3
	Scalability overhead [%] (as compared to single-layer coding)	10.0	10.8	12.6	15.2

Figs. 3 and 4 show that also the middle and enhancement layer bitstreams can be flexibly controlled by appropriate partitioning of the DCT non-zero coefficients.

5. CONCLUSIONS

With the same bitrate as by MPEG-2 nonscalable profile, the scalable coder proposed reaches almost the same quality. The bitrate overhead due to scalability is only about 5% - 25%. The three-layer codec proposed outperforms spatially scalable MPEG-2 or MPEG-4 codecs which generate bitrate overheads often exceeding 50% even for two-layer versions.

The paper proves that the coder proposed can be easily controlled in such a way that similar bitstreams are allocated to all three layers. Such bit allocation is very advantageous for practical applications.

Bitrate of the base layer coder is controlled using standard procedures well developed for single-layer coders. Three-dimensional spatio-temporal decimation allows the base layer bitrates of about 30 – 40 % of the total bitrate. The remaining bitstream, i.e. 60 – 70 % of the total bitstream can be partitioned between the other two layers. In particular, the middle layer bitrate is controllable in a relatively wide range as shown in Fig. 3. The ratio of bitrates related to the both 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. Fine granularity is also obtainable for the base layer because some nonzero DCT coefficients from the base layer can be also allocated into the middle layer using similar data partitioning

REFERENCES

[1] ISO/IEC IS 14496, Information Technology - Generic Coding of Audiovisual Objects.

[2] ISO/IEC IS 15444, Information Technology - JPEG 2000 Image Coding system.

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

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

[5] U. Benzler: Spatial scalable video coding using a combined subband-DCT approach, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, 2000, pp. 1080-1087.

[6] B.J. Kim, Z. Xiog, W. Pearlman: Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees, *IEEE Trans. Circ. Syst. Video Technology*, vol. 10, 2000, pp. 1374-1387.

[7] J. Tham, S. Ranganath, A. Kassim, Highly scalable wavelet –based video codec for very low bit-rate environment, *IEEE J. Select. Areas Commun*, vol. 16, 1998, pp. 12-27.

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

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

[10] M. Domański, A. Łuczak, S. Maćkowiak: Spatio-temporal scalability for MPEG video coding, *IEEE Trans. Circuits Systems for Video Technology*, vol. 10, 2000, pp. 1088-1093.

[11] M. Domański, S. Maćkowiak:, Modified MPEG-2 video coders with efficient multi-layer scalability, *Proc. IEEE Int. Conf. Image Proc.*, Tessaloniki, 2001, (in print).

[12] M. Domański, S. Maćkowiak: MPEG 2 – Based video coding with three-layer mixed scalability, *CAIP 2001 Proceedings*, Warszawa 2001, (in print).

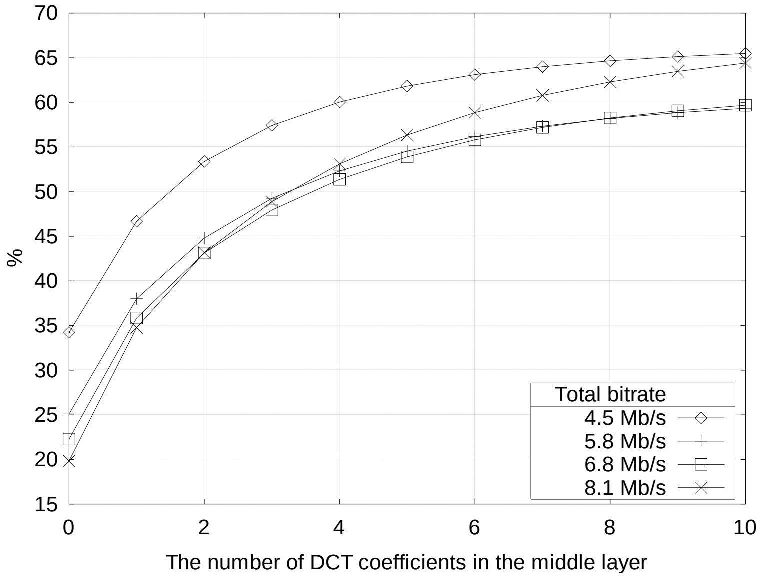

Fig. 3. Middle layer bitstream as percent of the total bitstream for test sequence *Fun*.

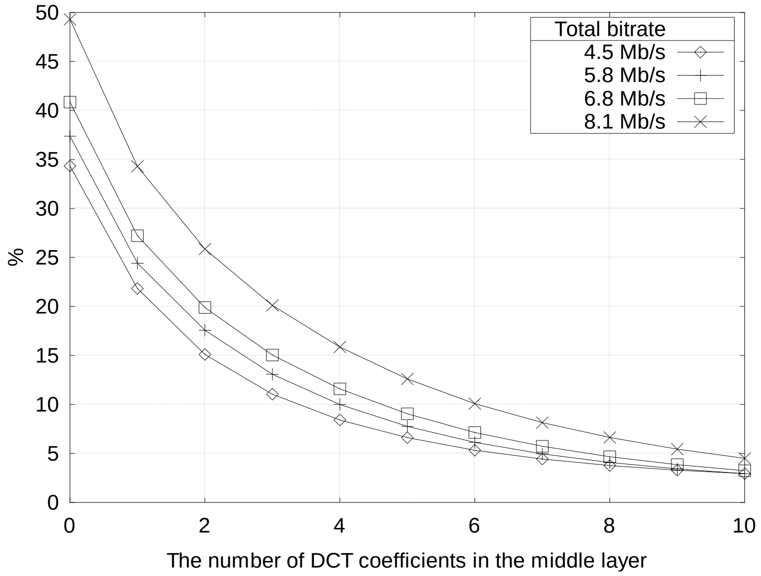

Fig. 4. Enhancement layer bitstream as percent of the total bitstream for test sequence *Fun*.