MODIFIED MPEG-2 VIDEO CODERS WITH EFFICIENT MULTI-LAYER SCALABILITY

Marek Domański, Sławomir Maćkowiak

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

e-mail: {domanski, smack }@et.put.poznan.pl

ABSTRACT

The paper describes a multi-layer video coder based on spatio-temporal scalability and data partitioning. The base layer represents video sequences with reduced spatial and temporal resolution. Decoding of a middle layer gives full resolution images but with lower quality as compared to those obtained from the enhancement layer also. The bitrate overhead measured relative to the single layer MPEG-2 bitstream varies about 10% - 25% for progressive television test sequences.. The coder is able to produce three layer-bitstreams with similar bitrates. The base layer is fully MPEG-2 compatible and the whole structure exhibits high level of compatibility with individual building blocks of MPEG-2 coders.

1. INTRODUCTION

There exist rapidly growing interests in video transmission through heterogeneous communication networks that are characterized by various maximum bitrates available for transmission. The available throughput depends on existing infrastructure that is very expensive to extend. There exist several examples, e.g. the service of video on demand using ADSL modems where available bitrate for an individual subscriber depends on the quality of the subscriber loop and the distance between subscriber and the exchange.

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 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). The respective functionality is called spatial, temporal or SNR scalability.

The example with video on demand on ADSL emphasizes the importance of scalability in the range of television broadcast bitrates, i.e. at bitrates of order of few Mbps. This is actual area of applications of the solutions discussed in the paper.

This paper deals with scalable multi-layer systems designed for bitrates of few Mbps. 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 all 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.

Among various proposals for spatially scalable coding of video, application of subband/wavelet decomposition should be considered as very promising [4-9]. Unfortunately, in most of such coders, it is difficult to allocate appropriate number of bits to the layers. A practical requirement is that the system produces layers with similar bitrates and this paper deals with such systems.

In order to meet this requirement for two-layer systems, 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 [10],
- Combination of spatial and temporal scalability called spatio-temporal scalability [11, 12]. Here the base layer represents a video sequence with reduced both temporal and spatial resolutions.

This paper deals with the latter approach that was considered in two versions hitherto. The first one exploited three-dimensional spatio-temporal decomposition [11]. The second version was based on partitioning of data related to B-frames [11, 12]. This approach as well as that based on combination space and SNR scalabilities were quit successful.

The paper deals with a novel version of spatio-temporal scalability based on partitioning of data related to B-frames where subband decomposition is exploited for the I-frames only [13,14]. The paper extends these concepts onto multi-layer systems.

2. MULTI-LAYER SCALABILITY

The concept is derived from the idea of two-layer coding [13,14]. The proposed coder can be described as a two-layer coder with spatio-temporal scalability where the enhancement layer is split into some further layers by use of data partitioning.

The coder will be described as a three-layer system but the concept can be extended onto the systems with more layers. The encoder will be presented for progressive video.

In the base layer, temporal resolution reduction is achieved by partitioning of the stream of B-frames: each second frame is not included into the base layer. Therefore there exist two types of B-frames:

• BE-frames which exist in the middle and the enhancement layers only,

• BR-frames which exist in all layers.

The I- and P-frame exist always in all 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.

The 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 other layers. These motion vectors MV are transmitted in the base layer.

The other part of the coder produces bitstreams for both middle and enhancement layer. It 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_m sent in the middle layer is four times that of the base layer.

2.1. I-frame coding

I-frames are split into four subbands using wavelet decomposition. The LL-subband (the subband of lowest frequencies) is sent with the base layer bitstream while the other three subbands are encoded for middle and enhancement layers.

2.2. P-frame coding

P-frames in the middle and enhancement layer encoder 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.

2.3. B-frame coding

Improved prediction is proposed for the BR-frames, which are the B-frames represented in all layers [13]. Each macroblock in a full-resolution BR-frame can be predicted from the following three reference frames (Fig. 1):

- previous reference frame (I- or P-frame from the enhancement layer),
- next reference frame (I- or P-frame from the enhancement layer),
- interpolated current reference frame (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 do not exist in the base layer are predicted using a modified MPEG-procedure. 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. 1) 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 [13].

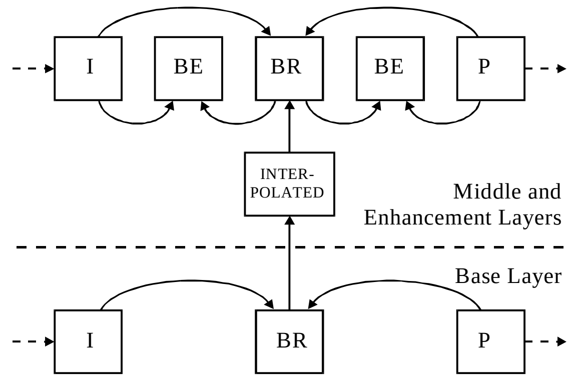


Figure 1. Prediction of B-frames.

2.4. Data partitioning

Second data partitioning is used to split the data produced by the enhancement encoder (Fig. 2) into middle and enhancement layers.

The middle layer consists of all header data and the enhancement motion vectors MV_m . The DCT coefficients are encoded as pairs (*run, level*) as described in the MPEG-2 standard. The stream of intra-DC coefficients as well as (*run, level*) pairs encoded by appropriate variable-length codes are partitioned into both layers in order to obtain requested bitrates in the middle and enhancement layers. For individual blocks, intra-DC and some beginning (*run, level*) pairs are allocated to the middle layer while the other (*run, level*) pairs are left to the enhancement layer. In order to preserve the MPEG-2 error resilience the slice headers are repeated in the enhancement layer.

For more than three layers, the goal of data partitioning would be production of more than two layers.

2.5. Coder structure

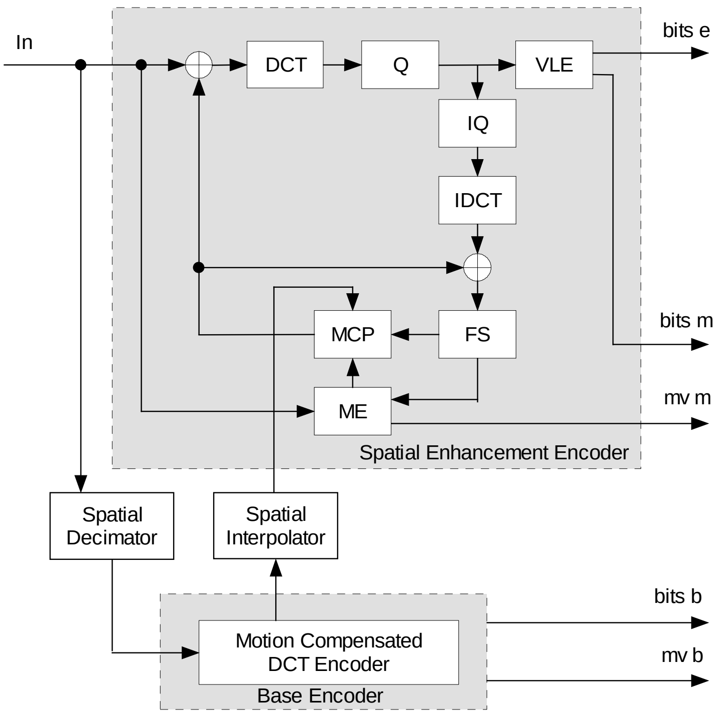
The coder is shown in Fig.2. It consist of a standard MPEG-2 base layer coder and MPEG-2-like middle/enhancement layer coder. The whole coder is only a slightly modified MPEG-2 scalable coder and consists of the blocks that are used in MPEG-2 coders.

3. EXPERIMENTAL RESULTS

The verification model has been written in C++ language and is currently available for progressive sequences with the input resolution defined by the standard digital television resolution. The software runs on PC-compatible workstations under Windows NT operating system.

The experiments have been made with progressive 720 × 576, 50 Hz, 4:2:0 test sequences (Fig. 3).

The experiments prove that drift present in the video sequences decoded without the highest enhancement layer is negligible. It is because of the influence from the base layer that is feeding the other layers with correct data.



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

Figure 2. The general structure of a three-layer coder (*bits b* and *mv b* – base layer, *bits m* and *mv m* – middle layer, *bits e* – enhancement layer).

Table 1. The experimental results for BT.601 progressive sequences.

		Funfair	*Flower Garden*	*Stefan*	*Cheer*	*Bus*
Single layer coder (MPEG-2)	Bitsream [Mb]	5,18	5,27	5,14	5,21	5,19
	Average PSNR [dB] for luminance	32,18	30,97	35,09	31,93	34,57
Proposed scalable coder	Base layer average PSNR [dB] for luminance	33,1	32,0	36,8	31,7	34,9
	Average PSNR [dB] for luminance recovered from both base and middle layers	29,2	29,1	32,3	29,5	30,9
	Average PSNR [dB] for luminance recovered from all three layers	32,2	30,9	35,1	32,0	34,5
	Base layer bitstream [Mb]	2,16	2,17	2,14	2,15	2,15
	Middle layer bitstream [Mb]	2,21	2,7	2,3	2,04	2,37
	Enhancement layer bitstream [Mb]	1,35	1,66	1,52	1,26	1,94
	Total bitstream [Mb]	5.72	6.53	5.96	5.45	6.46
	Base layer bitstream as percent of the total bitstream	37,8	33,2	35,9	39,5	33,3
	Middle layer bitstream as percent of the total bitstream	38.6	41.4	38.6	37.4	36.7
	Enhancement layer bitstream as percent of the total bitstream	23.6	25.4	25.5	23.1	30.0
	Scalability overhead [%]	10,4	23,9	15,9	4,6	24,5

Figure 3. A frame from the base, middle and enhancement layer of a sequence *Funfai*r (for the coding parameters given in Table 1).

4. CONCLUSIONS

In the base layer, the decoder receives a video sequence with reduced both temporal and spatial resolutions. The resolutions of pictures for middle and enhancement layers is full both in time and space but the quality of the pictures obtained from both base and middle layers is lower than those decoded from all three layers.

The coder is able to produce three bitstreams with similar bitrates. Such bit allocation is very advantageous for practical applications. 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 about 5% - 25%. The codec proposed outperforms spatially scalable MPEG-2 [1] or MPEG-4 [3] coders which generate bitrate overheads often exceeding 50% even for two-layer versions.

ACKNOWLEDGEMENT

The work has been supported by the Polish National Committee for Scientific Research under Grant 8 T11D 009 17.

REFERENCES

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

[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] 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.

[6] F. Bosveld, *Hierarchical video compression using SBC*, Ph.D. dissertation, Delft Univ. of Technology, Delft 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] 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] 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.

[11] 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, 1998.

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

[13] A. Łuczak, S. Maćkowiak, M. Domański, Spatio-temporal scalability using modified MPEG-2 predictive video coding”, in *Signal Processing X: Theories and Applications*, Tampere, 2000, pp. 961-964.

[14] M. Domański, A. Łuczak, S. Maćkowiak, On improving MPEG spatial scalability, *Proc. Int. Conf. Image Proc.*, Vancouver, 2000, vol. 2, pp. 848-851.