MPEG 2 – Based Video Coding
with Three-Layer Mixed Scalability

Marek Domański, Sławomir Maćkowiak

Poznań University of Technology,
Institute of Electronics and Telecommunication,
Piotrowo 3A, 60-965, Poznań, Poland
Phone: +48 61 66 52 762
Fax: +48 61 66 52 572
e-mail: {domanski, smack }@et.put.poznan.pl

Abstract. The paper describes a three-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 5% - 25% for progressive television test sequences. 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. The paper reports experimental results that prove useful properties of the coder proposed.

1 Introduction

There exists 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. 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 some layers are decodable into video sequences with reduced spatial resolution, temporal resolution or signal-to-noise ratio (SNR). This functionality is called spatial, temporal or SNR scalability, respectively.

For practical reasons, the number of layers is often limited to three or even two layers as in most profiles of MPEG 2 [1,2]. This papers deals with a proposal of a video coding system with three-layer scalability (cf. Fig. 1).

The paper is focused on the functionality of spatial scalability that is already provided by the MPEG-2 [1,2] and MPEG-4 [3] video compression standards. Unfortunately, respective standard implementations are mostly related to unacceptably high bitrate overheads mostly about 50% to 70% of the single-layer bitrate. Among other proposals for spatial scalability, those based on application of subband/wavelet decomposition are the most popular [4-9] but they suffer from the problems with bit allocation and interframe coding.
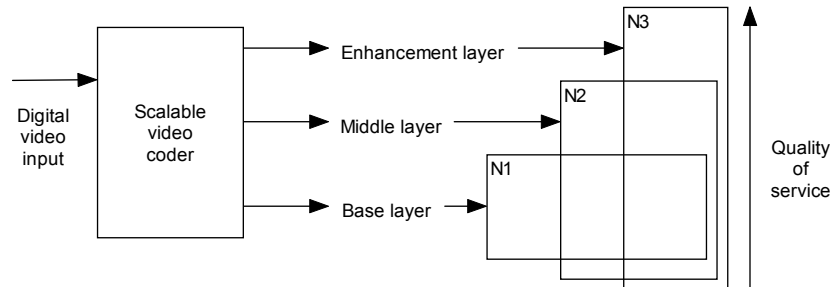
Fig. 1. Three-layer scalable video transmission in heterogeneous network that consists of subnetworks N1, N2 and N3 accessible via links with various Quality of Service

The goal of the work is to achieve possible low bitrate overhead, i.e. the total bitrate related to all three layers of a scalable bitstream should be as close as possible to the bitrate of the respective single-layer coder. The paper deals with scalable video coders that compress digital television signals to bitrates of order of few megabits per second. These bitrates are suitable, e.g. for television broadcasting, the service of video on demand and video transmission using ADSL technology. Moreover it is required that the system produces layers with similar bitrates.

In order to meet the latter 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 based on a combination of spatial and SNR scalability [10] or spatial and temporal scalability [11, 12]. This paper deals with the latter approach of spatio-temporal scalability 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 three-layer systems.

The important assumption made for this work 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.

2 Spatio-temporal scalability with B-frame data partitioning

For the sake of simplicity, spatio-temporal scalability will be reviewed for the simplest case of two-layer systems, i.e. systems that produce the base layer bitstream and the enhancement layer bitstream. Base layer bitstream allows for restoration of video with all resolutions (horizontal, vertical, temporal) halved as compared to the

input video sequence. Enhancement layer bitstream can be additionally used for restoration of the full-quality video.

Let us assume that the number of B-frames between two consecutive I- or P-frames is odd. Temporal resolution reduction can be achieved by partitioning of the stream of B-frames: each second frame is included into the enhancement layer only. Therefore there exist 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. Motion vectors are transmitted for the base layer. Another motion vectors MV_e are estimated for the enhancement layer. In the enhancement layer, difference values (MV_e-MV_b) are transmitted.

Improved prediction is used 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 three reference frames: previous reference frame (I- or P-frame), next reference frame (I- or P-frame), current reference frame (BR-frame). The improvement on standard MPEG-2 prediction within a single layer consists in another decision strategy. The best prediction/interpolation is chosen from all three possible reference frames: previous, future and interpolated.

The 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.

3 Three-layer scalable system

In the previous section a two-layer scalable coder has been described. Experimental data [12-14] obtained by the authors prove that the enhancement layer bitrate is mostly larger than that of the base layer. Therefore a reasonable three-layer system can be designed by use of data partitioning of the enhancement layer, i.e. the enhancement layer data are partitioned into the middle layer data and the new enhancement layer data according to the rule described by Table 1.

The enhancement layer repeats slice headers because each layer has to be error resilient. The repeated slice headers allow the decoder to resynchronize the enhancement layer bitstream after a transmission error.

The middle layer contains N_m first pairs (*run, level*) for individual blocks. The control parameter N_m influences bitrate allocation between the middle and enhancement layers.

The bitrates in the individual layers can be controlled individually to some extent. The experience of the authors shows that all the three bitrates can be set to similar values for a wide range of the overall bitrates of the scalable coder. The coder structure is shown in Fig. 2.

Table 1 Data partitioning in a three-layer system

Middle layer	Enhancement layer of the three-layer system
1. Headers 2. Motion vectors 3. Low frequency DCT coefficients	1. Slice headers 2. Remaining DCT coefficients
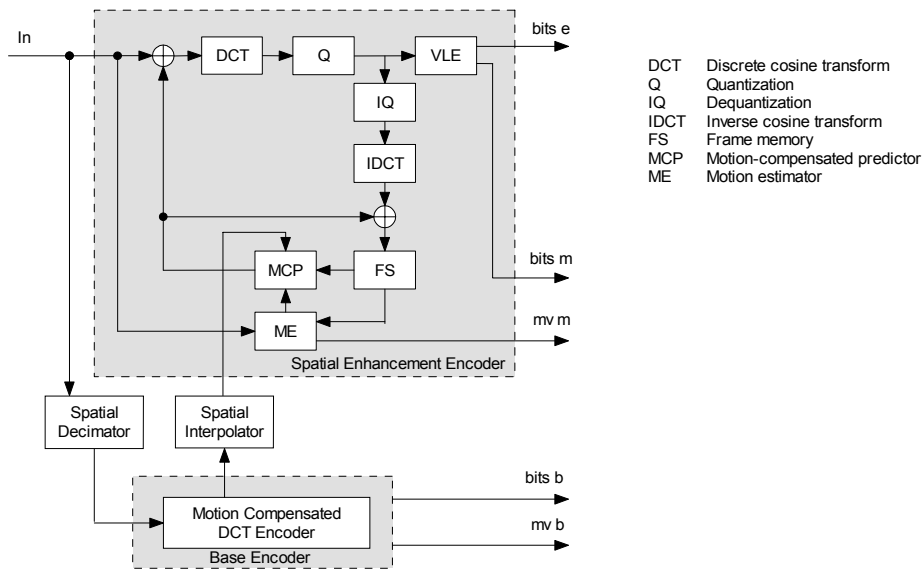
Fig. 2. The general structure of a three-layer coder (*bits b* and *mv b* – base layer, *bits m* and *mv m* – middle layer, *bits e* – enhancement layer)

4 Experimental results and conclusions

The verification model has been prepared as software written in C++ language. The coder is implemented for processing of progressive 4:2:0 720 × 576 sequences with 50 frames per second. The base layer coder is standard MPEG 2 coder that processes video in the SIF format but both the middle layer and the enhancement layer are in the full television resolution.

For the sake of brevity, the results are given for two video sequences only. The Table 2 summarizes the results for bitrates in the range of 4 – 7 Mbps. The results for a nonscalable MPEG 2 MP @ ML coder are given for comparison. Figures 3 and 4 include rate-distortion plots for all layers. Fig. 5 shows frames reconstructed from all three layers of quality.

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

References

1. ISO/IEC IS 13818, Generic Coding of Moving Pictures and Associated Audio Information
2. Haskell, B., Puri, A., Netravali, A.: Digital Video: An Introduction to MPEG-2. Chapman & Hall, New York (1997)
3. ISO/IEC IS 14496, Generic Coding of Audiovisual Objects
4. Tsunashima, T., Stampleman, J., Bove V.: A Scalable Motion -Compensated Subband Image Coder. IEEE Trans. Communication, **42** (1994) 1894-1901
5. Gharavi, H., Ng, W.Y.: H.263 Compatible Video Coding and Transmission. In: Proc. First International Workshop on Wireless Image/Video Communication, Loughborough (1996) 115-120
6. Bosveld, F.: Hierarchical Video Compression Using SBC, Ph.D. dissertation, Delft Univ. of Technology, Delft (1996)
7. Senbel, S., Abdel-Wahab, H.: Scalable and Robust Image Compression Using Quadtrees. Signal Processing: Image Communication **14** (1999) 425-441
8. Shen, K., Delp, E.: Wavelet Based Rate Scalable Video Compression. IEEE Trans. Circ. Syst. Video Technology **9** (1999) 109-122
9. Chang, P.-C., Lu, T.-T.: A Scalable Video Compression Technique Based on Wavelet and MPEG Coding. In: Proc. Int. Conf. on Consumer Electronics (1999) 372-373
10. Benzler, U.: Spatial Scalable Video Coding Using a Combined Subband-DCT Approach. IEEE Trans. Circ. Syst. Video Technology **10** (2000) 1080-1087
11. Domański, M., Łuczak, A., Maćkowiak, S., Świerczyński, R.: Hybrid Coding of Video with Spatio-Temporal Scalability Using Subband Decomposition. In: Signal Processing IX: Theories and Applications, Rhodes (1998) 53-56
12. Domański, M., Łuczak, A., Maćkowiak, S.: Spatio-Temporal Scalability for MPEG Video Coding. IEEE Trans. Circ. Syst. Video Technology **10** (2000) 1088-1093
13. Łuczak, A., Maćkowiak, S., Domański, M.: Spatio-Temporal Scalability Using Modified MPEG-2 Predictive Video Coding. In: Signal Processing X: Theories and Applications, Tampere (2000) 961-964
14. Domański, M., Łuczak, A., Maćkowiak, S.: On Improving MPEG Spatial Scalability, In: Proc. Int. Conf. Image Proc., Vancouver (2000) 848-851

Table 2 The experimental results for BT.601 progressive test sequences.

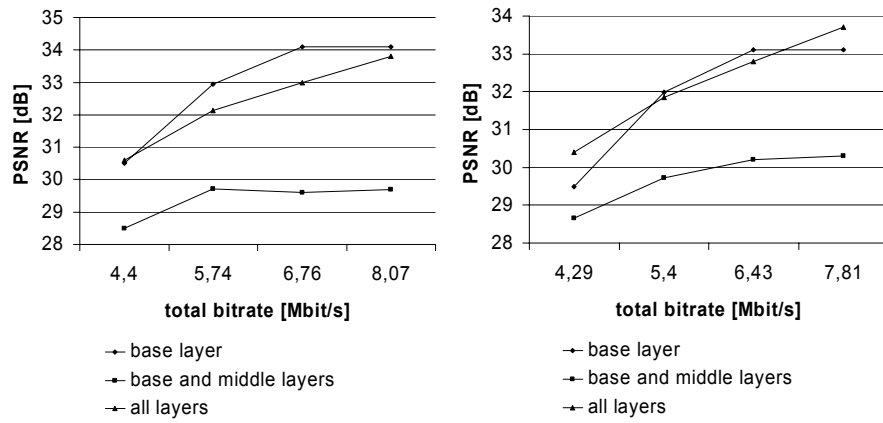| Test sequence	Fun Fair				Cheer			
Single layer coder (MPEG-2)								
Bitstream [Mb/s]	4.0	5.18	6.0	7.0	4.0	5.21	6.0	7.0
Average PSNR [dB] for luminance	30.6	32.2	33.0	33.8	30.6	31.9	32.8	33.7
Proposed scalable coder								
Base layer average PSNR [dB] for luminance	30.5	33.0	34.1	34.1	29.5	32.0	33.1	33.1
Average PSNR [dB] for luminance recovered from both base and middle layers	28.5	29.7	29.6	29.7	28.7	29.7	30.2	30.3
Average PSNR [dB] for luminance recovered from all three layers	30.6	32.1	33.0	33.8	30.4	31.9	32.8	33.7
Base layer bitstream [Mbit/s]	1.40	2.16	2.50	2.50	1.40	2.15	2.50	2.50
Middle layer bitstream [Mbit/s]	2.00	2.18	2.42	2.80	1.97	2.03	2.28	2.67
Enhancement layer bitstream [Mbit/s]	1.00	1.40	1.84	2.77	0.92	1.22	1.65	2.64
Base layer bitstream as percent of the total bitstream	31.8	37.6	37.0	31.0	32.6	39.8	38.9	32.0
Middle layer bitstream as percent of the total bitstream	45.5	38.0	35.8	34.7	45.9	37.6	35.5	34.2
Enhancement layer bitstream as percent of the total bitstream	22.7	24.4	27.2	34.3	21.5	22.6	25.6	33.8
Scalability overhead [%] (as compared to single-layer coding)	10.0	10.8	12.6	15.2	7.25	2.6	7.2	11.6

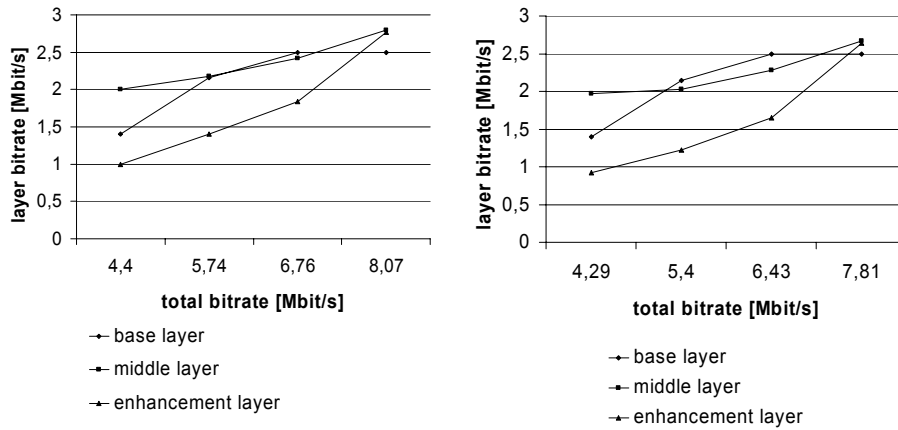**Fig. 3.** PSNR for test sequences *Funfair* and *Cheer*

Fig. 4. Layer bitrates for test sequences *Funfair* and *Cheer*

Fig. 5. A frame from the base, middle and enhancement layer of a test sequence *Funfair* (for the coding parameters given in Table 3)