Spatio-Temporal Scalable Video Codecs with MPEG-Compatible Base Layer

Marek Domanski[1*], Adam Luczak[*],
Slawomir Mackowiak[*], Roger Swierczynski[*], Ulrich Benzler[◆]

[*]Poznan University of Technology, Institute of Electronics and Telecommunications,
Poznan, Poland
[◆]University of Hannover, Institute of T. Telecommunications and Information Processing,
Hannover, Germany

Abstract - A new technique of scalable coding is proposed for progressive video with bitrates of order of megabites per second. The goal is to improve spatial scalability of MPEG-2 by introducing spatio-temporal scalability. The technique proposed needs less coding overhead than in MPEG-2 spatially scalable scheme and an enhancement layer bitstream with its bitrate not less than the bitrate in a base layer. The solution proposed in the paper is based on both temporal and spatial resolution reduction performed for data transmitted in a base layer. The temporal resolution reduction is obtained by placing each second frame (B-frame) in the enhancement layer. The enhancement layer includes also high-frequency spatial subbands from other frames. The base layer is fully MPEG-2 compatible.

1. INTRODUCTION

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. Therefore low-resolution terminals are able to decode only the base layer bitstream in order to display low-resolution pictures. In error-resilient video transmission, base layer packets are often better protected against transmission errors while the protection of the enhancement layer is lower. A receiver is able to reproduce at least low-resolution pictures if quality of service decreases.

The functionality of spatial scalability is already provided by the MPEG-2 video coding standard [1,2]. Unfortunately, its implementation based on pyramid decomposition (cf. Fig. 1) is not satisfactory in many applications. By many test sequences, the total bitstream is not much smaller than sum of bitstreams obtained for simulcast transmission with two different resolutions.

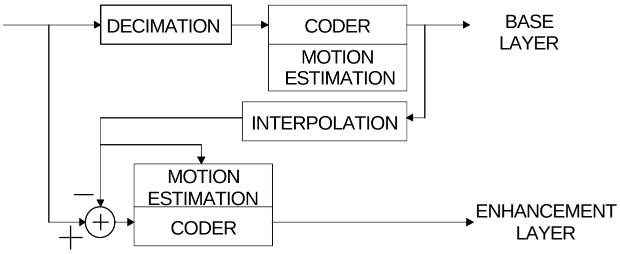

Fig. 1. *The basic idea of spatial scalability as proposed in MPEG standards.*

The goal of this paper is to propose alternative techniques to achieve spatial scalability for SDTV or HDTV resolutions. The assumption is that the low-resolution base layer is fully MPEG-2 compatible.

It is also assumed that similar bitstreams in the base layer and the enhancement layer will be obtainable in the scheme proposed.

---

[1] The corresponding author: Marek Domañski, ul. Piotrowo 3a, 60-965 Poznañ, Poland,
Phone.: +48 61 8782762; fax: +48 61 8782572, e-mail: domanski@et.put.poznan.pl

The work has been supported by the National Committee for Scientific Research under Grant 8 T11D 007 11.

2. SUBBAND DECOMPOSITION

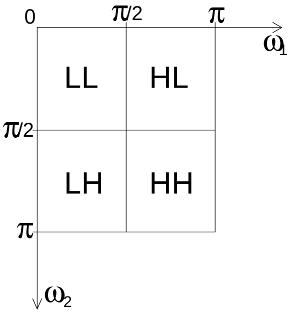

Fig. 2. *Subband decomposition.*

There were many attempts to improve the scheme of spatial scalability by application of subband decomposition. The idea is to split each image into four spatial subbands (Fig. 2). The subband of lowest frequencies constitutes a base layer while the other three subbands are jointly transmitted in an enhancement layer.

Embedding of subband decomposition into a motion-compensated coder leads to in-band or out-band compensation performed on individual subbands or whole image, respectively. Some experimental results show that the latter is more efficient [3].

Unfortunately, direct application of the above scheme does not allow to control flexibly the bitstreams of the base and enhancement layers. Therefore hybrid spatio-temporal scalability is proposed in order to obtain further data reduction in the base layer [4,5].

3. SPATIO-TEMPORAL SCALABILITY

The term of spatio-temporal scalability is proposed for video compression systems where a base layer corresponds to pictures with reduced both spatial and temporal resolution. An enhancement layer is used to transmit the information needed for restoration of the full spatial and temporal resolution.

The two basic approaches related to spatio-temporal scalability are related to three-dimensional subband analysis as well as to temporal resolution reduction by use of partitioning of the sequence of B-frames [4,5]. Here, the latter approach is considered.

For sake of simplicity, interlaced video is not considered in the paper.

Reduction of temporal resolution is obtained by removal of each second frame from a video sequence. The technique employs data structures already designed for standard MPEG-2 coding. It is assumed that groups of pictures (GOPs) consist of even number of frames. Moreover, it is assumed that each second frame is a B-frame that can be removed from a sequence without affecting decodability of the remaining frames.

Reduction of spatial resolution is obtained by use of subband decomposition. Proper design of filter bank results in negligible spatial aliasing in the LL subband which is used as the base layer. Unfortunately the technique does not provide any means to suppress temporal aliasing. The effects of temporal aliasing are similar as those related to frame skipping in hybrid coders.

Standard order of frames in the base and enhancement layers is as in Table 1.

Table 1. *GOP structure in both layers*

Base layer (only subband LL)	Enhancement layer
I	I (without LL subband)
skipped	B
B	B (without LL subband)
skipped	B
P	P (without LL subband)
skipped	B
B	B (without LL subband)
skipped	B
P	P (without LL subband)
skipped	B
B	B (without LL subband)
skipped	B
P	P (without LL subband)
skipped	B
B	B (without LL subband)
skipped	B

The base layer is the low-quality channel therefore it is reasonable to perform here more rough quantization than in the enhancement layer. On the other hand, quality of the subband LL is strongly related to the quality of the full-sized picture. Low quality of the LL subband restricts the full-sized picture quality to a relatively low level despite of the amount of information in the remaining subbands. Therefore it is substantial to transmit additional information ΔLL in the enhancement layer. This information is used to improve quality of the subband LL when used to synthetize a full-sized image in the enhancement layer.

4. SCALABLE CODER

The fundamental assumption which restricts the structure of a coder is that base layer must be MPEG-2-compatible.

The structure of the coder is shown in Fig. 3.

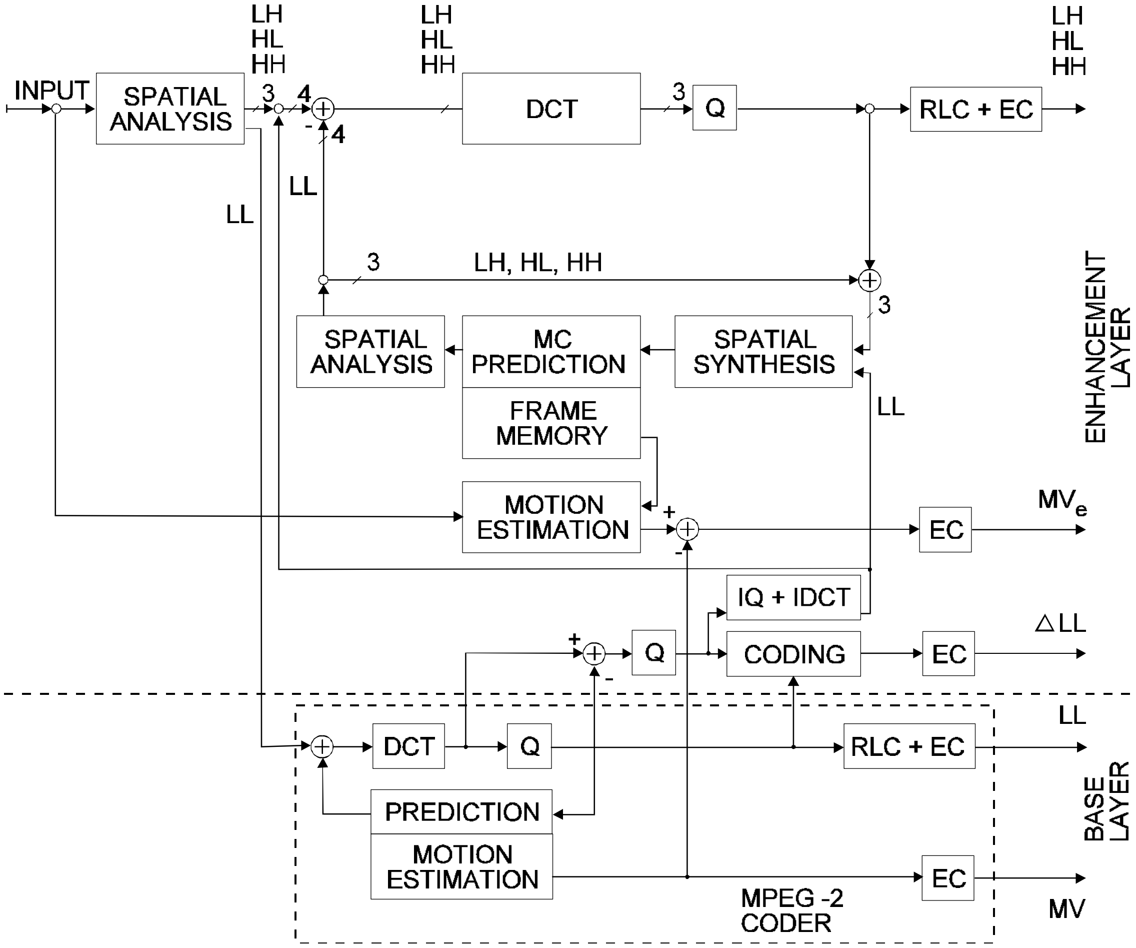

Fig. 3. *Block diagram of the coder.*

Base layer coder is implemented as a motion-compensated hybrid MPEG-2 coder. This coder supplies the enhancement layer coder with three data streams:

- DCT coefficients from LL subband,
- quantized DCT coefficients from LL subband,
- motion vectors.

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, HH).

Therefore there are two subband analysis stages and one subband synthesis stage in the coder.

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.

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.

5. IMPLEMENTATION AND 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 14 000 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 test sequences. Therefore the base layer is in the SIF format. The experiments have been done with 4:2:0 sequences.

The analysis and synthesis filter banks were Johnston 12th order FIR QMFs [6] and Daubechies (9,3) wavelets [7].

The preliminary experimental results for progressive 720 × 576, 50 Hz test sequences (*”Mobile & calendar*”, *”Flower garden*”, *”Basketball”* and *”Funfair”*) are given in Table 3. The first two test sequences were obtained via deinterlacing of SDTV interlaced sequences.

Table 2. *Experimental results.*

	Test sequence						
	Flower Garden		Mobile Calendar	Basketball		Funfair	
Single layer (MPEG-2) bitstream [Mbps]	4.90	7.57	7.89	4.67	6.99	4.67	6.15
Average luminance PSNR [dB]	29.6	32.6	30.8	29.9	32.1	31.3	32.9
Scalable bitstream [Mbps]	5.22	8.66	10.70	5.02	9.26	4.97	8.45
Average luminance PSNR [dB]	29.6	32.4	29.9	29.6	31.1	31.0	32.3
Base layer bitstream [Mbps]	2.13	3.01	3.35	2.07	3.12	2.02	3.17
Base layer bitstream [%]	40.8	34.8	31.3	40.3	33.7	40.6	37.5
Bitstream overhead [%]	6.5	14.4	35.6	7.5	32.5	6.4	37.4

The first goal, i.e. base layer bitstream not exceeding that of enhancement layer has been reached for all bitrates and video test sequences tested.

Nevertheless, coding efficiency measured by the bitrate overhead for scalable coding as compared to MPEG-2 nonscalable coder was satisfactory not for all bitrate/sequence pairs. The experiments with lower bitrates mostly resulted in the overhead of order of 6–15%. In such cases, the coder proposed strongly outperforms the spatially scalable MPEG-2 coder.

The problem to solved is how to control efficiently the parameters of this quit sophisticated coder.

It must be noted that the results were obtained for standard (not optimized) quantizers and standard encoding of the ΔLL stream. The coder is still not sufficiently tuned. Prospective tuning of its parameters and proper choice of quantization and coding tables is related to further data reduction. First of all, the ΔLL stream must be reduced. In the less successful experiments cited above, ΔLL is often between 130 kbit and 340 kbit per frame. For B-frames it is most data used. On the other hand, the successful experiments were related to smaller bitrates of the ΔLL stream. Therefore the goal of current work is to reduce ΔLL stream in the B-frames and possibly in the P-frames.

Further data reduction is obtainable by optimization of the quantization tables in the LH, HL and HH subbands.

Our observation is that HH subband could be skipped for most frames. It would decrease bitrate by sparing the header bits that must be sent even for very few bits allocated for the subband.

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*, Chapman & Hall, 1997.

[3] F. Bosveld, *Hierarchical video compression using SBC*, Doctoral dissertation, Delft University of Technology, Delft 1996.

[4] M. Domanski, A. Luczak, S. Mackowiak, R. Swierczynski, Hybrid coding of video with spatio-temporal scalability using subband decomposition, *Signal Processing IX: Theories and Applications*, Rhodes, 1998, pp. 53-56.

[5] M. Domanski, A. Luczak, S. Mackowiak, R. Swierczynski, Hybrid coding of video with spatio-temporal scalability using subband decomposition, *Proceedings of SPIE* vol. 3653, 1999, pp. 1018-1025.

[6] J. Johnston, A filter family designed for use in quadrature mirror filter banks, Proc. IEEE Int. Conf. Acoustics Speech Signal Proc., 1980, pp. 291-294.

[7] I. Daubechies, *Ten lectures on wavelets*, SIAM 1992.