MPEG-2 COMPATIBLE HIERARCHICAL VIDEO CODERS WITH B-FRAME DATA PARTITIONING

Marek Domanski, Adam Luczak, Slawomir Mackowiak
Poznañ University of Technology, Institute of Electronics and Telecommunications,
Piotrowo 3A, 60-965, Poznañ, Poland,
Ee-mail:{ domanski, aluczak, smack}@et.put.poznan.pl

ABSTRACT

A new technique of spatio-temporal scalability coding is considered for video with bitrates of order of megabits per second. Both temporal and spatial resolutions are reduced 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. Improved prediction of B-frames yields very high efficiency of the technique proposed.

1. INTRODUCTION

The paper deals with multiresolution video coding where a low resolution image is decodable from base layer data while a full resolution image can be obtained from data of both base and enhancement layers. Such compression techniques are of great interest recently because of development of communication networks with different transmission bitrates, e.g. [1,2].

Multiresolution video coders called also 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.

Application of scalable coding allows low-resolution terminals to decode only the base layer bitstream in order to display low-resolution pictures. Growing interest in hierarchical coding is related to development of high-resolution digital television (HDTV) in parallel to standard digital television (SDTV).

Moreover scalable transmission is advantageous in error-prone environments. 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 [3,4]. Unfortunately, its implementation based on pyramid decomposition 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.

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.

2. SPATIO-TEMPORAL SCALABILITY

There were many attempts to improve the scalable multiresolution coding of video. Among various proposals, application of subband decomposition should be considered as very promising [e.g. 5,6]. The idea is to split each image into four spatial subbands. The subband of lowest frequencies constitutes a base layer while the other three subbands are jointly transmitted in an enhancement layer (Fig. 1). Nevertheless this approach often leads to allocation of much higher bitrates to a base layer than to an enhancement layer which is disadvantageous for practical applications.

In order to avoid the above mentioned problem spatio-temporal scalability is proposed [7,8]. Here, a base layer corresponds to the bitstream of the pictures with reduced both spatial and temporal frequencies. Therefore less data is transmitted in the base layer. An enhancement layer is used to transmit the information needed for restoration of the full spatial and temporal resolution. It is easy to get the base layer bitstream equal or even less than that in the enhancement layer.

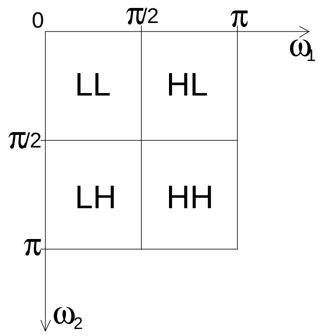

Fig. 1. Subband decomposition

2. SPATIO-TEMPORAL SCALABILITY WITH B-FRAME DATA PARTITIONING

Temporal resolution reduction is achieved by partitioning of the stream of B-frames: each second frame is included into the enhancement layer. We assume that the number of B-frames between two consecutive I- or P-frames is odd. Therefore each second frame included in the enhancement layer is a B-frame.

Spatial resolution reduction is achieved by subband decomposition as shown in Fig. 1.

This idea is summarized in Tables 1 and 2. Table 2 explains an exemplary structure of a group of pictures (GOP) being a data unit known from MPEG-2.

Table 1. Layers

Base layer	Enhancement layer
Subband LL from each even frame.	Subbands LL, LH, HL, HH from each even frame and each odd frame (B-frame).

Table 2. Frames of different types in both layers

Base:	I		B		P		B		P		B		P
Enhancement:	I	B	B	B	P	B	B	B	P	B	B	B	P

3. SYSTEM STRUCTURE

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.

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.

The overall structure of the coder is shown in Fig. 2.

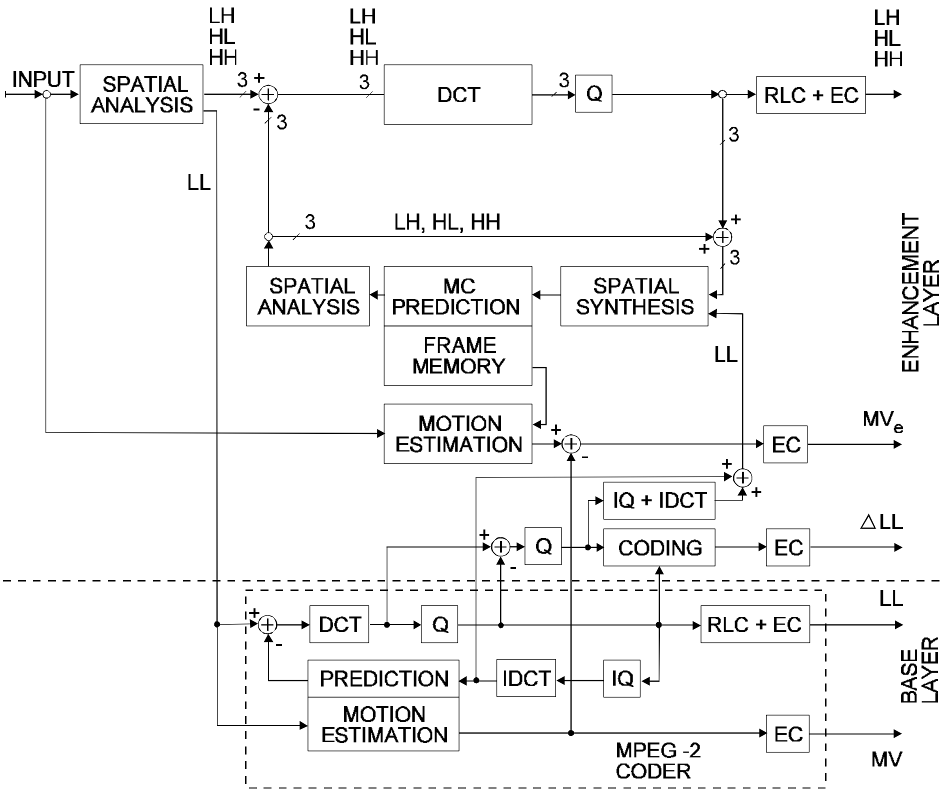

Fig. 2. Block diagram of the coder.

4. IMPROVED B-FRAME PREDICTION

In order to improve coding efficiency, B-frame prediction (in the enhancement layer) uses also base layer B-frames as shown in Fig. 3. Experimental results with SDTV test sequences prove that this arrangement improves significantly encoding of B-frames.

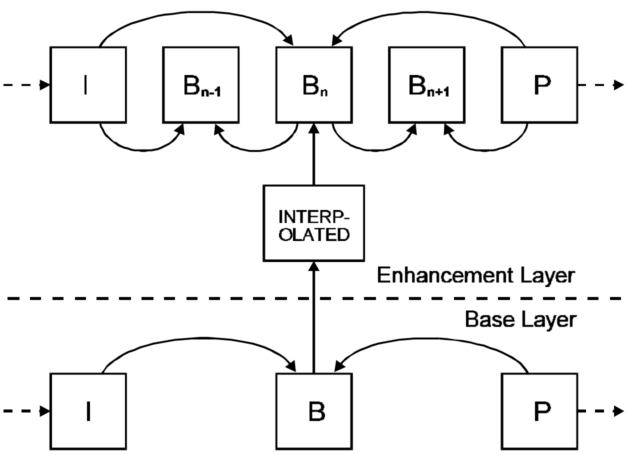

Fig. 3. Improved prediction of B-frames.

5. IMPLEMENTATION AND EXPERIMENTAL RESULTS

The codec was implemented as software in C++ language. The most important feature of this software 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 preliminary experimental results for progressive 720×576, 50 Hz test sequences (”*Flower garden*”, and ”*Funfair”*) are given in Table 3.

Table 3. Experimental results

		Test sequence			
		Flower Garden		*Funfair*	
Single Layer (MPEG-2)	Bitsream [Mb]	5.24	8.76	5.03	8.48
	Average luminance PSNR [dB]	30.63	33.75	31.38	34.07
Proposed scalable coder	Bitstream [Mb]	5.34	8.74	5.04	8.48
	Average luminance PSNR [dB]	30.44	33.41	30.81	34.16
	Base layer bitstream [Mb]	2.21	3.00	2.11	2.70
	Base layer bitstream [%]	41	34	41	31

6. CONCLUSIONS

The experimental results from Table 3 prove high efficiency of the coder. With the same bitrate as by MPEG-2 nonscalable profile, the scalable coder proposed reaches almost the same quality. The codec proposed significantly outperforms spatially scalable solutions standardized in the MPEG-2 recommendation [3].

ACKNOWLEDGEMENT

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

REFERENCES

[1] Shen, E. Delp, wavelet based rate scalable video compression, IEEE Trans. Circuits Syst. Video Technology, vol. 9, 1999, pp. 109-122.

[2] Senbel, H. Abdel-Wahab, Scalable and robust image compression using quadtrees, Image Communication, vol. 14, 1999, pp. 425-441.

[3] ISO/IEC International Standard 13818, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information.*

[4] B. Haskell, A. Puri, A. N. Netravali, *Digital Video: an Introduction to MPEG-2,* Chapman & Hall, 1997.

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

[6] H.Gharavi, W.Y.Ng, *H.263 Compatible Video Coding and Transmission*, First Int. Workshop on Wireless Image/Video Communication, Loughbo-rough 1996, pp. 115-120.

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

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