Marek Domanski
Adam Luczak
Slawomir Mackowiak
Roger Swierczynski
Poznan University of Technology
Institute of Electronics
and Telecommunications
ul. Piotrowo 3a, 60-965 Poznan

MPEG-Based Scalable Video Codec with 3-D Subband Decomposition

Keywords: MPEG-2, spatio-temporal scaleablility, 3-D subband decomposition

ABSTRACT

The MPEG-2 video coding standard provides the functionality of spatial scalability in an inefficient way. Therefore solutions based on subband decomposition gained a lot of interest very recently. In particular, this paper deals with scalable (hierarchical) coding of video at bitrates in the range of about 2-6 Mbps. The solution proposed in the paper is based on both temporal and spatial resolution reduction performed for data transmitted in a base layer. The basic idea is to employ three-dimensional subband decomposition. The assumption is that the base layer is encoded fully MPEG-2 compatible. Bitstreams related to high spatial frequencies are encoded using a technique that exploits mutual dependencies between low- and high- spatial-frequency subbands. The technique is implemented as a C++ program with about 12 980 lines of code.

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. An important feature of scalable bitstreams is that a 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. The functionality of spatial scalability is also important for error-resilient video transmission where base layer packets are well protected against transmission errors or packet losses 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 is inefficient because the number of pixels to be encoded is increased as compared to a nonscalable scheme. Therefore there were many attempts to improve the scheme of spatial scalability by application of subband decomposition, e.g. [3-5]. 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. 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. Here, 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 assumption is that the base layer is fully MPEG-2 compatible.

In order to encode efficiently the subbands related to high spatial frequencies, a technique that exploits mutual dependencies between spatial subbands is proposed. The idea is to transmit only the portions of the high frequency images that correspond to ”active” portions in the low-frequency subband. ”Active” portions of the low-frequency subband are defined as those where differences between neighboring pixels exceed a predefined threshold used to control the compression ratio.

The codec is implemented as a C++ program with 12 980 lines of code.

2. THREE-DIMENSIONAL SUBBAND ANALYSIS

The input video sequence is analyzed in an three-dimensional (3-D) separable filter bank, i.e. there are three consecutive steps of analysis: temporal, horizontal and vertical (Fig. 1).

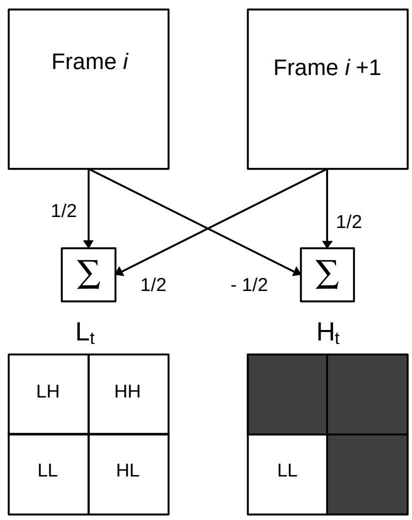

Fig. 1. Three-dimensional subband analysis.

For temporal analysis, very simple linear-phase two-tap filters are used similarly as in pure three-dimensional subband coding [7,8]

$$H(z) = 0.5\,(1 \pm z^{-1}),$$

where "+" and "-" correspond to low- and high-pass filters, respectively. This filter bank has a very simple implementation and exhibits small group delay (half of sampling period) resulting in small system response times which are very critical for many applications in interactive multimedia.

Temporal analysis results in two subbands L_t and H_t which are partitioned into four spatial subbands (LL, LH, HL and HH) each. For spatial analysis, both horizontal and vertical, separable nonrecursive half-band linear-phase filters in polyphase implementation are used. Linear phase is assumed here because motion vectors estimated in the LL subband of the L_t temporal subband are used for motion compensation in the LL subband of the H_t temporal subband.

The three-dimensional analysis results in eight spatio-temporal subbands. Three high-spatial-frequency subbands (LH, HL and HH) in the high-temporal-frequency subband H_t are discarded as they correspond to the information being less relevant for the human visual system. Therefore five subbands are encoded (see Fig. 1):

- In a base layer **-** the spatial subband LL of the temporal subband L_t .
- The enhancement layer includes the spatial subbands LH, HL and HH from the temporal subband L_t and the spatial subband LL of the temporal subband H_t.

The base layer is produced by an MPEG-2 motion compensated coder.

3. CODER STRUCTURE

The coder structure has been chosen after some experiments with various variants of coding algorithm (Fig. 2). The experiments have been made for progressive format of video, and interlaced video is not considered in this paper.

As mentioned above, three high-spatial-frequency subbands (LH, HL and HH) in the high-temporal-frequency subband H_t are discarded as they correspond to the information being less relevant for the human visual system. Although it reduces PSNR, e.g. in the intraframe mode to about 32-33 dB, it has small influence on subjective quality of the decoded video.

Two subbands are encoded using MPEG-compatible DCT-based coders. The base layer is produced by an MPEG-2-compatible coder which processes the subband LL from the temporal subband L_t. Motion estimation, the most computationally demanding task is performed in this subband. Quarter-pel accuracy is the goal of the motion estimation. The motion vectors obtained are used also for the LL subband from the temporal subband H_t which is encoded using an MPEG-2-compatible motion-compensated coder. Nevertheless this coder does not include a motion estimator.

Motion compensation in the LL subband from the temporal subband H_t is not very efficient. The experiments made resulted in a conclusion that the bitrate reduction caused by motion compensation usually does not exceed 20%.

As mentioned above, the LL/L_t subband is encoded using an MPEG-2 coder. In order to obtain high compression ratio for the base layer, rough quantization of the DCT coefficients is employed, e.g. with the quantization coefficient of about 8-16. Such quantization would decrease the quality of the full resolution image too much. Therefore some corrections for non-zero valued DCT coefficients are additionally transmitted in the enhancement layer. These corrections are encoded using Huffman codes and their locations are defined by the locations of the non-zero coefficients from the base layer.

The experiments prove that motion compensation in the three high-spatial-frequency subbands (LH, HL and HH) in the low-temporal-frequency subband L_t is inefficient. This corollary is similar to that from [6]. Therefore these subbands are encoded without motion compensation. In order to exploit mutual dependencies between subbands, an original technique previously proposed for still images [9] is used. The pixels which do not correspond to the ”active” portions of the LL/L_t subband are discarded. The remaining pixels are quantized by a nonlinear quantizer and encoded using DPCM and variable-length coding.

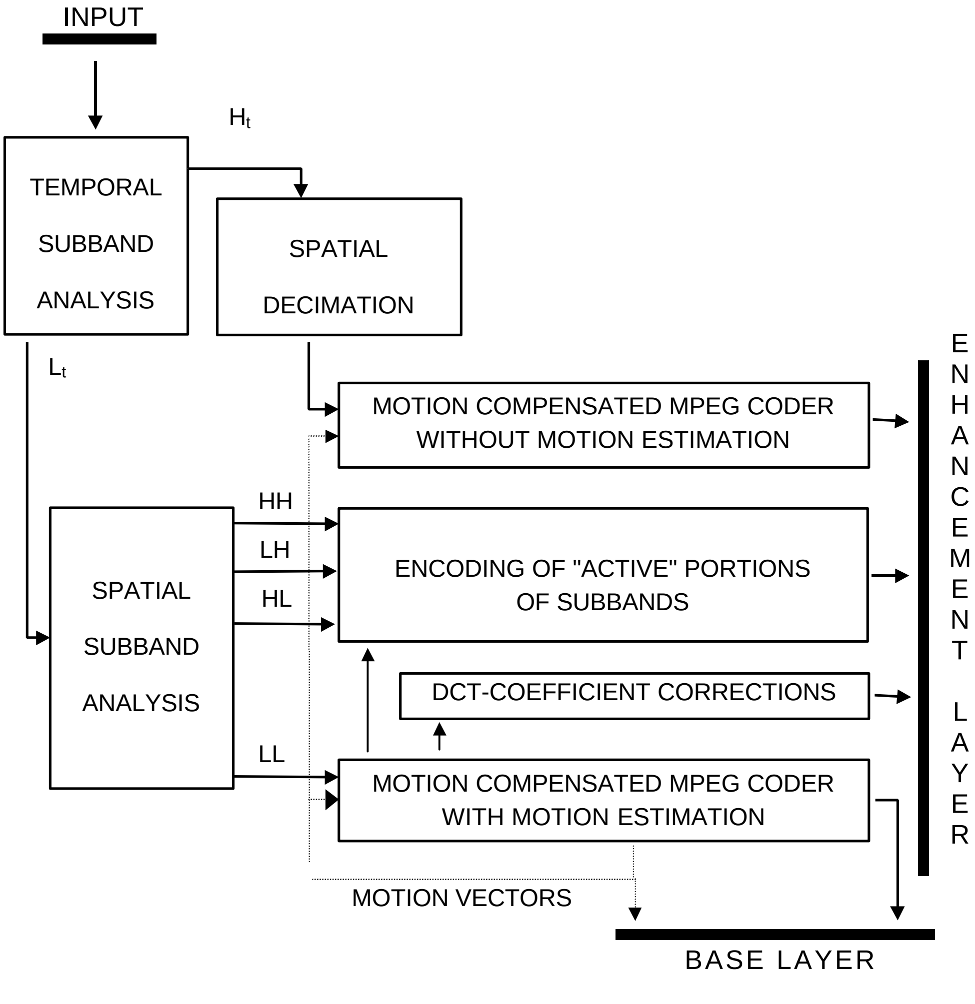

Fig. 2. Coder structure.

4. SOFTWARE AND EXPERIMENTS

For the sake of brevity, only the basic features of the codec are described. In fact, the coding algorithm is very sophisticated and its implementation needs a lot of work. The software prepared is enormous. In order to achieve flexibility in changing parameters standard MPEG-2 software [10] has not been used but even an MPEG-2 coder has been newly implemented.

The purpose of the experiments is to estimate the best coder structure and its properties. Therefore the software is written in C++ language and is not optimized for run time. The most important feature is its flexibility allowing tests of different variants of coding algorithm. Currently the program includes 12 980 lines of code. The software runs on Sun 20 workstations under Solaris operational system.

Using the software mentioned above, the described schemes have been examined with the standard test sequences. Their luminance format is progressive BT.601 [11], i.e. 720×576 pixels and 25 frames per second. The chrominance subsampling scheme is 4:2:0. It means that the base layer provides the SIF frames.

The preliminary experimental results for the test sequences ”*Mobile & calendar*” and ”*Flower garden*” are shown in Table 1.

Table 1: Preliminary experimental results.

Sequences	Bitstream Mbps	Base layer as percentage of total	PSNR for full resolution [dB]			PSNR for base layer [dB]		
			Y	Cr	Cb	Y	Cr	Cb
Flower garden	3.0	37.09	27.5	31.8	33.8	31.2	33.8	35.6
	3.5	37.95	28.2	32.4	34.2	32.3	34.7	36.2
	4.0	39.95	28.9	33.1	34.7	33.4	35.6	36.9
Mobile & calendar	3.3	35.46	26.9	31.8	32.8	29.4	32.8	33.5
	3.6	38.92	27.6	32.4	33.5	30.5	33.6	34.4
	4.0	43.32	28.3	33.2	34.2	31.8	34.6	35.3

The promising results show that the bitstream of the base layer is less than half of the total bitstream. The software needs some optimization and is currently on the final stage of development. Nevertheless the obtained rate-distortion function is already promising.

5. CONCLUSIONS

The paper describes a new scalable video coder. A new feature is application of three-dimensional subband decomposition in a scalable MPEG-compatible coder.

The structure of the coder has been chosen according to the results of preliminary experiments with test video sequences.

Enormous work has been already done by preparation the software that implements the video codec. This software is currently being tested but some promising preliminary experimental results have been already obtained.

Further work will include experiments which should estimate the properties of the coder. Moreover the coder needs optimization and tuning by adjustment of its parameters.

The goal of the work is to propose a structure which potentially could be proposed for standardization as a new MPEG-2 profile.

ACKNOWLEDGMENTS

The authors express their sincere thanks to Prof. H.G. Musmann and Mr. U. Benzler from Hannover University in Hannover (Germany) for their help, advice and fruitful discussions.

The work has been supported by the Grant KBN 8 T11D 007 11.

REFERENCES

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

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

[3] O. Tsunashima, J. Stampleman, V. Bove., *A scalable motion -compensated subband image coder*, IEEE Trans. on Communication, vol. 42, 1994, pp. 1894-1901.

[4] U. Benzler, *Scalable Multiresolution Video Coding Using Sub-Band Decomposition*, First Int. Workshop on Wireless Image/Video Communication, Loughborough 1996, pp. 109-114.

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

[6] Gharavi, *Subband coding of video signals*, chapter 6 in *Subband Image Coding*, edited by J. W. Woods, Kluwer Academic Publishers 1991

[7] J.Ohm, *Three-dimensional subband coding with motion compensation*, IEEE Trans. Image Proc., vol.3, 1994, pp.559-571.

[8] M. Domanski, R. Swierczynski, *3-D subband coding of video using recursive filter banks*, Signal Processing VIII: Theories and Applications, Trieste 1996, pp. 1359-1362.

[9] M. Domanski, R. Swierczynski, *Subband coding of images using hierarchical quantization*, Signal Processing VII: Theories and Applications, Edinbourgh 1994, pp. 1218-1221.

[10] MPEG Software Simulation Group, *MPEG-2 coder, ver. 1.2*, 1996.

[11] Recommendation ITU-R BT.601-4, *Encoding parameters of digital television for studios*, Geneva 1994.