HYBRID CODING OF VIDEO WITH SPATIO-TEMPORAL SCALABILITY USING SUBBAND DECOMPOSITION

Marek Domański*, Adam Łuczak, Sławomir Maćkowiak, Roger Świerczyński
* *IEEE and EURASIP member*

Poznań University of Technology,
Institute of Electronics and Telecommunications,
Piotrowo 3A, 60-965, Poznań, POLAND
Tel.: +48 61 8782762; fax: +48 61 8782572
e-mail: {domanski, aluczak, smack, roger}@et.put.poznan.pl

ABSTRACT

The paper deals with scalable (hierarchical) coding of video at bitrates in the range of about 2-5 Mbps. A practical goal is to obtain comparable bitrates in a base layer and an enhancement layer. The solution proposed in the paper is based on both temporal and spatial resolution reduction performed for data transmitted in a base layer. Two principal variants are presented: systems with three-dimensional filter banks for spatio-temporal analysis and systems where some B-frames are allocated to an enhancement layer but subband analysis is purely spatial. The assumption is that a base layer is fully MPEG-2 compatible. Bitstreams related to high spatial frequencies are encoded either using DCT-based MPEG-like coding or a technique that exploits mutual dependencies between low- and high-spatial-frequency subchannels.

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 because of their prospective applications in error-resilient video coding and multi-standard video transmission.

The functionality of spatial scalability is provided by the MPEG-2 video coding standard [1]. Unfortunately, its implementation based on pyramid decomposition is inefficient. Therefore 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. 3,4]. 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 the bitstream of the pictures with reduced both spatial and temporal frequencies. An enhancement layer is used to transmit the information needed for restoration of the full spatial and temporal resolution. The assumption is that a base layer is fully MPEG-2 compatible.

There are two basic variants of the spatio-temporal scalable system considered in the paper.

2 SYSTEMS WITH THREE-DIMENSIONAL SUBBAND ANALYSIS

The input video sequence is analyzed by temporal, horizontal and vertical filter banks (Fig.1). For temporal analysis, very simple linear-phase two-tap filters are used similarly as in pure three-dimensional subband coding [8,9]

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

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 applications. It enables perfect reconstruction in very practical DFT arrangements.

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 of motion estimation and compensation.

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 irrelevant for the human visual system. Therefore five subbands are encoded (see Fig.2):

- 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. Motion vectors obtained for the base layer (subband LL/L_t) are used also for the other four subbands from the enhancement layer.

There are several possibile methods to encode the enhancement layer. DCT-based motion-compensated MPEG-like coding is one of possible solutions.

Motion vectors are supplied to all four subbands of the enhancement layer. Nevertheless the gain obtained using motion compensation in the subbands of L_t - LH, HL and first of all in HH is poor. A better but more sophisticated solution consists in spatial synthesis of subbands and joint motion compensation for all subbands.

Quantization tables for the LL subbands in L_t and H_t are similar to standard MPEG-2 tables. Nevertheless for higher base layer compression ratios, it is reasonable to include DCT-coefficient updates into the enhancement layer. Otherwise rough quantization of the DCT coefficients in the base layer would prevent the system to reach better coding fidelity even by fine quantization in the enhancement layer.

The LH, HL and HH subbands (from L_t) need flat quantization tables.

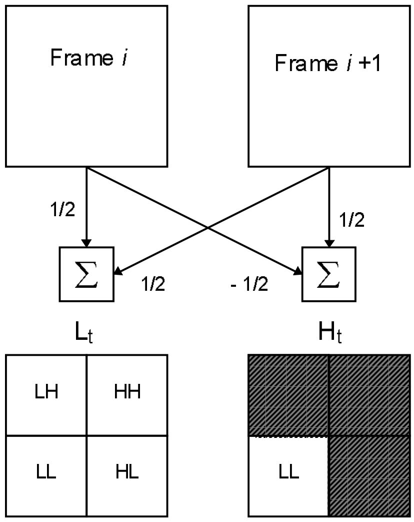

Figure 1: Three-dimensional subband analysis.

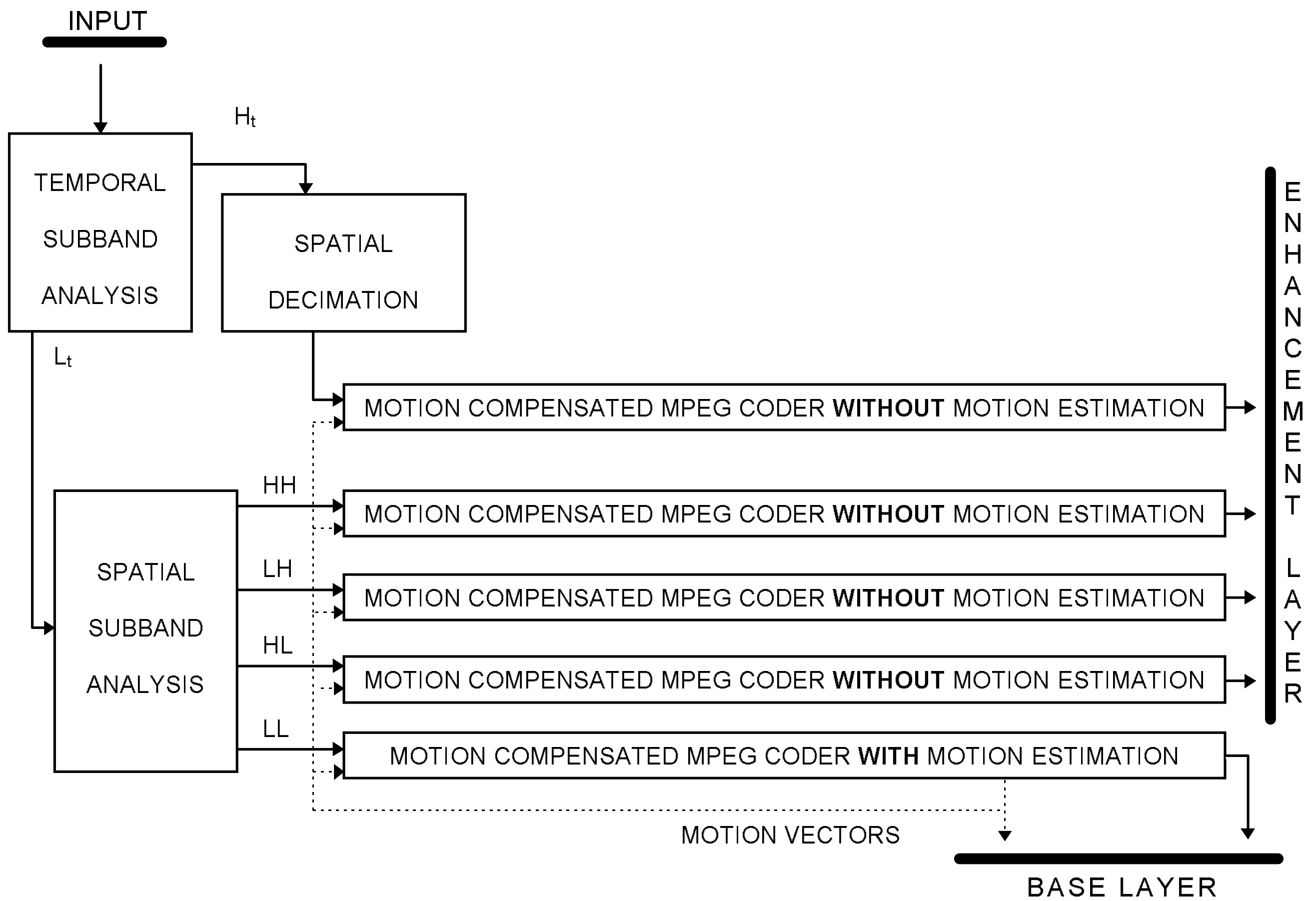

Figure 2: Scalable system with three-dimensional subband analysis.

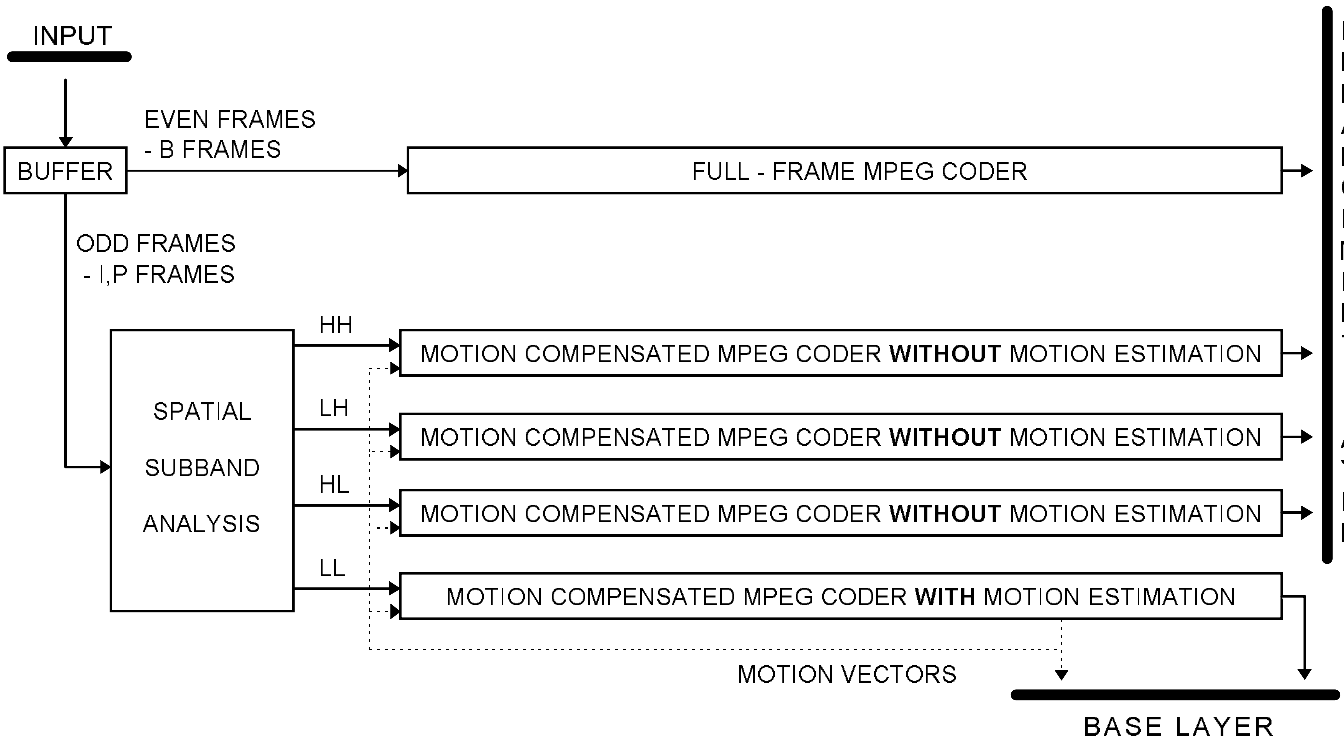

Figure 3: Second variant of a scalable coder.

3 INCORPORATION OF B-FRAMES INTO AN ENHANCEMENT LAYER

In the design of an alternative scalable system, it is assumed that every second input frame is coded as the B-frame and is never used for other frames prediction. Thus these frames need not to be sent in the base layer. The enhancement layer includes (see Fig.3):

- each even frame being a B-frame,
- the LH, HL, HH spatial subbands from the odd frames.

The base layer includes the subband LL from half of frames.

The spatial subbands as well as the B-frames are coded in the standard hybrid DCT-based coders. Motion vectors obtained for the LL subband are used for motion compensation of other spatial subbands. Following the suggestion of [5] the intraframe coding of the HH subband is considered as an option using spatial vector quantization.

Special quantization tables are calculated for individual subbands.

4 CODING EXPLOITING INTER-SUBBAND CORRELATION

In both above described systems, a video sequence at input to the spatial analysis filter bank is a sequence of pictures similar to those in the source video sequence. Usually pictures showing natural scenes exhibit very high cross-correlation between sub-images corresponding to different subbands of the spatial spectrum. Therefore a technique originally successfully applied for still images [10] can be adopted for efficient encoding of the subbands HL, LH and HH (Fig. 4).

This technique exploits an observation that images of natural scenes exhibit power spectra which decay more or less uniformly for increasing frequencies. Therefore, the portions of the high-frequency sub-images where signal values are substantial correspond usually to the portions of the low-frequency sub-image where some changes occur. Most of the portions of the high-frequency sub-images exhibit only small signal values. Therefore a quantizer with a dead-zone produces a lot of zeros. We need not to transmit all of them and we can use the low-frequency sub-image to identify the positions of the "active" portions of the high-frequency sub-images.

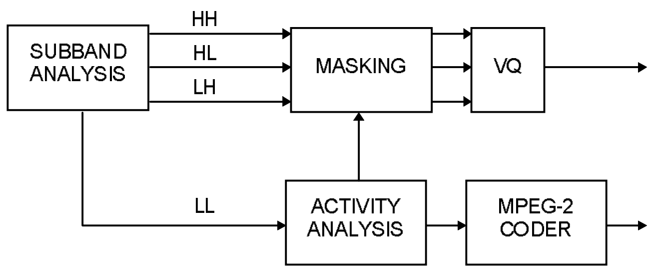

Figure 4: Coding exploiting inter-subband correlation.

"Active" portions of the low-frequency sub-image correspond to details and textures in the original image. We define some activity measures in the LL subband. Only the portions of the sub-images in the high-frequency subbands that correspond to the "active" portions in the LL subband are encoded. Other portions of the high-frequency sub-images are skipped in order to reduce the amount of data.

The remaining data in the LH, HL, HH subbands are encoded using geometrical vector quantization [11]. There are separate codebooks for each subband. For each block 4 x 4 pixels a vector number and its quantized amplification coefficient is transmitted.

This technique is currently under development and testing.

5 EXPERIMENTAL RESULTS

The described schemes have been examined with the standard CCIR-601 progressive test sequences. 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. These results have been obtained for the first variant of the system where subband decomposition in time and space is performed. DCT-based MPEG-like coding was employed in all subbands as shown in Fig.2.

6 CONCLUSIONS

The compression software is currently being developed. Therefore preliminary experimental results are described above for the first system variant only.

Preliminary results show that the proposed systems can be considered as prospective solutions for the scalable video coding in tv-applications.

A corollary from first experiments is that DCT-based coding is inefficient for high-frequency subbands. The technique exploiting inter-subband correlation and vector quantization is much more promising.

ACKNOWLEDGEMENTS

The authors express their sincere thanks to Prof. H.G. Musmann and Mr. U. Benzler from the University of 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] Tsunashima, J. Stampleman, V. Bove., *A scalable motion -compensated subband image coder*, IEEE Trans. on Communication, vol. 42, 1994, pp. 1894-1901.

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

[4] 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.

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

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

[7] M. Uz, M. Vetterli, D. J. LeGall, *Interpolative multiresolution coding of advanced television with compatible subchannels*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 1, no. 1, March 1991.

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

[9] Domański, R. Świerczyński, 3-D subband coding of video using recursive filter banks, Signal Processing VIII: Theories and Applications, Trieste 1996, pp. 1359-1362.

[10] Domański, R. Świerczyński, Subband coding of images using hierarchical quantization, Signal Processing VII: Theories and Applications, Edinbourgh 1994, pp. 1218-1221.

[11] Ch. Podlichuk, N. Jayant, N. Farvardin. Three-dimensional subband coding of video, IEEE Trans. Image Proc., vol.4, 1995, pp.125-284.

Table 1: Results for a system that exploits a three-dimensional analysis.

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