Scalable multi-resolution video coding using a combined subband-DCT approach

Ulrich Benzler

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung

Universität Hannover

Appelstr. 9A, D-30167 Hannover

Phone: +49-511-762-5047, Fax: +49-511-762-5052

Email: benzler@tnt.uni-hannover.de

Abstract

A combined subband-DCT approach for scalable multi-resolution video coding is presented. Therefore, each image is decomposed into 4 subbands by analysis filtering, whereat the low frequency subband represents the base layer, and together with the high frequency subbands the enhancement layer is reconstructed by synthesis filtering. The low frequency subband and the three high frequency subbands are separately coded using DCT-based techniques similar to MPEG-2. Further, for adjusting the bit rate distribution between base and enhancement layer, SNR scalability for the base layer is used. Compared with non-scalable MPEG-2 coding, the additional overhead for scalability is only 5-9% of the total bit rate, whereas the MPEG-2 Spatial Scalable Profile (SSP) leads to an overhead of 66-70%.

1. Introduction

Scalable multi-resolution video coding requires that parts of the encoded and transmitted data can be decoded separately by a base layer decoder to reconstruct low resolution images. In combination with unequal error protection it can be used to accomplish graceful degradation in the case of transmission errors in a mobile communication application. This functionality is also of interest for information servers in order to allow access with different channel capacities.

In MPEG-2 and MPEG-4 this functionality is provided by the spatial scalable profiles (SSP) [1][2] where a pyramid coding scheme is used. Low resolution images are derived by low pass filtering and subsampling the original video signal. The low resolution images are encoded separately, and the resulting bitstream represents the base layer information of the scalable system. After decoding the base layer, the reconstructed low resolution images are up-converted and additionally used to generate a prediction signal for the high resolution images (see Figure 1). The resulting prediction error is encoded in the enhancement layer at a sample rate which is equal to that of the original video input signal. Therefore, the total number of encoded samples exceeds the number of video input samples by the number of samples encoded in the base layer. This disadvantage of an increased sampling rate cannot fully be compensated by providing an additional prediction signal for the high resolution images. Thus, an overall coding efficiency can be observed which is only slightly better than 'simulcast' where the low resolution images and the high resolution images are encoded independently [3].

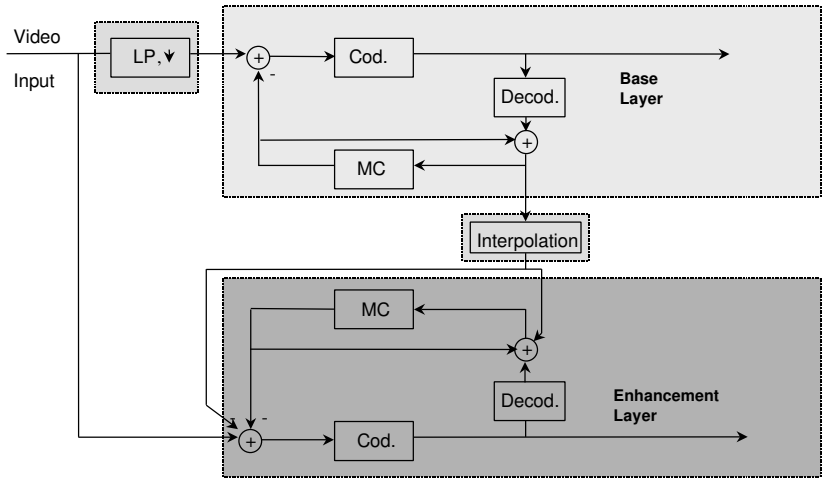

Figure 1 : MPEG spatial scalable coding scheme

2. Scalable multi-resolution subband-DCT video coding

In the proposed subband-DCT scheme the input video signal is decomposed by a 4-band analysis filter bank (AF) resulting in one low frequency and three high frequency bands (see Figure 2). Similar to SSP, the low frequency subband is encoded separately in the base layer for providing low resolution images. But unlike SSP, the subband decomposition reduces the sample rate in the enhancement layer which now carries only the three high frequency bands each sampled at one fourth of the original rate. High resolution images are reconstructed by a synthesis filter bank (SF) using both the low frequency band reconstructed from the base layer bit stream and the three high frequency bands reconstructed from the enhancement layer bit stream. The motion compensated prediction signal for the three high frequency bands of the current image is generated by applying motion compensating prediction (MC) to a previously reconstructed high resolution image and subsequent analysis filtering (see Figure 2). Due to the critically sampled subband decomposition, the total number of encoded samples is equal to the number of video input samples. Hence, the disadvantage of an increased sample rate is completely avoided.

Furthermore, both the base and the enhancement layer use the same motion information, which reduces the necessary side information. This is possible by the use of quarter pel motion compensation in the base layer, which was successfully proposed for the forthcoming MPEG-4 standard [2]

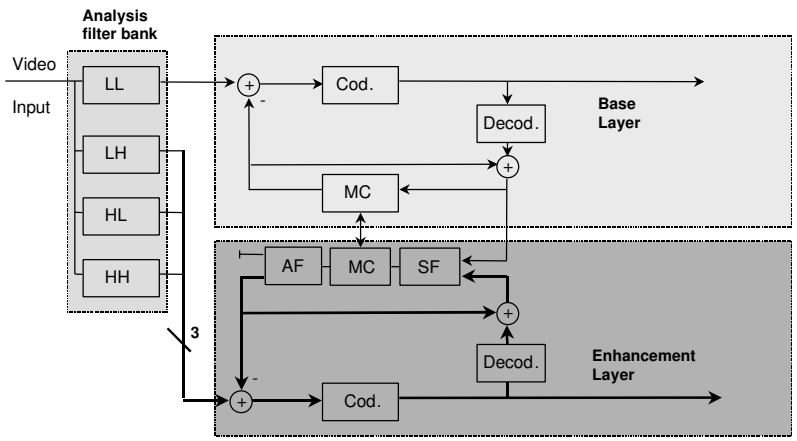

Figure 2 : Combined scalable subband-DCT coding

Experimental results comparing the proposed scheme to non-scalable and scalable MPEG-2 coding show that the combined scalable subband-DCT encoder needs only 5 to 9 % more bit rate than a non-scalable MPEG-2 encoder whereas the spatial scalable MPEG-2 encoder increases the bit rate by 66 to 70 % (see Table 1).

PSNR	30 dB		32 dB	
	Mbit/s	Relatively [%]	Mbit/s	relatively [%]
Non-scalable MPEG-2	4,14	100,0	6,26	100,0
MPEG-2 SSP	7,06	170,1	10,43	166,6
Combined scalable subband-DCT	4,44	107,4	6,61	105,5

Table 1 : Experimental results for constant quality coding (ITU-R 601 test sequence flower garden)

3. SNR Scalability for the base layer

In order to provide the functionalities outlined above, channel coding for error protection and possible scalable modulation require that the base layer bit rate does not exceed 50% of the total bit rate [4].

An analysis of the bit allocation between base and enhancement layer of the developed coding system shows that base layer coding needs approx. 75 % of the total bit rate.

An analysis of the non-scalable MPEG-2 encoder shows that depending on the test sequence 65 to 80% of the overall bit rate is used for coding the frequencies equivalent to those carried in the base layer of the scalable system.

This is due to the low pass characteristic of the Modulation Transfer Function (MTF) of the human visual system. The MTF allows a larger quantisation error at high spatial frequencies, leading to a larger number of zero-amplitude coefficients which reduces the demand in bit rate for the high frequency bands. In addition the base layer also carries the common side information for motion vectors, prediction information etc., which increases the base layer bit rate even more.

In order to overcome this bit allocation problem, data of the low frequency information need to be transmitted in the enhancement layer. The solution presented here is the so-called 'SNR scalability' where a coarsely quantised low frequency subband is transmitted in the base layer, using only 50 % of the total bit rate. Additional data for fine quantisation of the low frequency subband is transmitted in the enhancement layer, together with the three high frequency bands. In this work 'embedded quantisers' [5] in combination with bit plane coding [6] are used for achieving this goal. The modified block diagram is shown in Figure 3.

4. Simulation results

The developed coding scheme including the SNR scalability for the base layer is compared with non-scalable MPEG-2 coding at both base and enhancement layer resolution. The total bit rate for the scalable coder is 6 Mbit/s, the base and the enhancement layer bit rate are 3 Mbit/s each. Two ITU-R 601 test sequences (91 frames from Basketball and Bus) are used.

For both base and enhancement layer the additional cost for scalability is less than 10% of the total bit rate compared with non-scalable coding at the same PSNR, or less than 0.5 dB PSNR degradation at the same bit rate. For the base layer scalable and non-scalable coding show the same performance (see Figure 4).

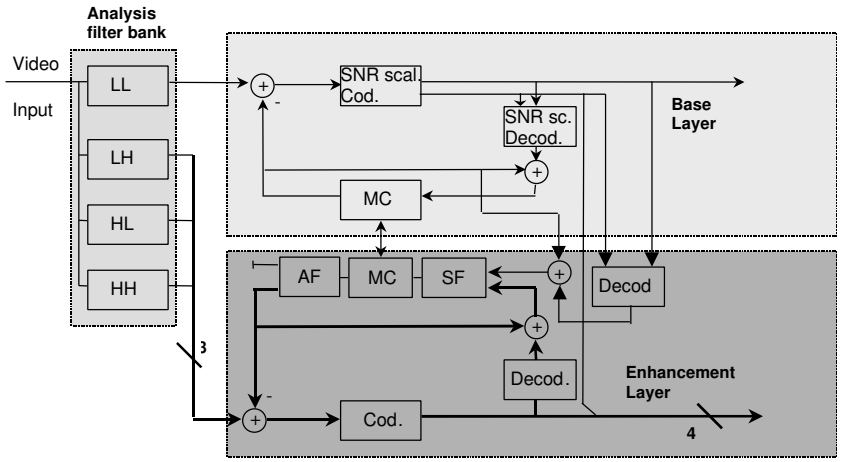

Figure 3 : Scalable subband-DCT-coding with
SNR Scalability for the base layer

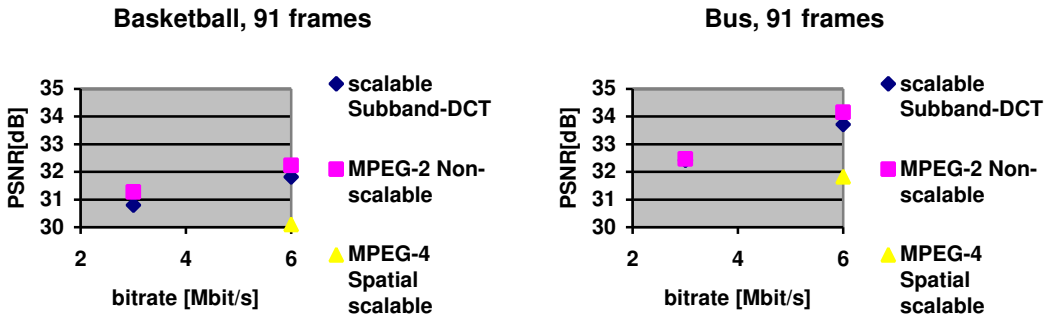

Figure 4 : Simulation results for scalable subband-DCT coding compared with non-scalable MPEG-2 coding (3 Mbit/s for low resolution, 6 Mbit/s for high resolution)

5. Conclusions

A scalable video coding scheme is presented. Combined scalable subband-DCT coding provides different spatial resolutions at different bit rates with a small overhead of less than 10% of the total bit rate or less than 0.5 dB decrease in PSNR compared with non-scalable MPEG-2 coding. The proposed SNR scalability for the base layer allows a finer adjustment of the bit rate allocation for the base and the enhancement layer bitstreams, resulting in the same bit rate for base and enhancement layer .

6. References

[1] MPEG-2 : ISO/IEC, “Revised Text for ITU-T Recommendation H.262 | ISO/IEC 13818-2 : Information technology – Generic coding of moving pictures and associated audio information : video”, March 1995

[2] MPEG-4 : ISO/IEC, “Working Draft 6.0 for International Standard of ISO/IEC 14496-2 Version 2 : Information Technology – Coding of Audio-Visual Object : Visual”. ISO/IEC JTC1/SC29/WG11 N2553, December 1998

[3] N. Wells, P. Tudor, “Standardization of scalable coding schemes”, Proc. IEEE ISCAS 94, Tutorials, pp 121-130, May 1994

[4] Seeger, A., “ Hierarchical channel coding for Rayleigh and Rice fading”, IEEE Globecom, Communication Theory Mini Conference, Phoenix, November 1997, pp. 208-212

[5] Bosveld, F., “Hierarchical Video Compression using SBC”. Delft: Universiteit Delft, Faculteit der Elektrotechniek, Diss., pp 80-84, 1996

[6] W. Li et. al., “Bit-plane coding of DCT coefficients”. Doc. ISO/IEC JTC1/SC29/WG11 MPEG97/2691, October 1997