ON IMPROVING MPEG SPATIAL SCALABILITY

Marek Domański, Adam Łuczak, Sławomir Maćkowiak

Poznań University of Technology, Institute of Electronics and Telecommunications,
Piotrowo 3A, 60-965, Poznań, Poland

ABSTRACT

The paper describes video codecs that provide the functionality of spatio-temporal scalability that enables construction of highly scalable video codecs. The coder exploits wavelet decomposition for intraframe coding and combined temporal prediction and spatial interpolation for P- and B-frames. In a two-layer system, the bitrate overhead measured relative to the single layer MPEG-2 bitstream varies about 10% for progressive television test sequences. The base layer bitstream constitutes about 40% of the overall bitstream. In multilayer systems, the bitrate ratio of consecutive layers is about 1:2.5. The base layer encoder is fully compatible with the MPEG-2 video coding standard. The paper comprises experimental results obtained for a two-layer system for progressive video sequence compression.

1. INTRODUCTION

Spatially scalable or hierarchical video coders produce a set bitstreams: a base layer bitstream which represents low resolution pictures and the bitstreams of enhancement layers which provide additional data needed for reproduction of pictures with full resolution. The base layer bitstream can be decoded independently from enhancement layers. Additional decoding of some enhancement layers provides data necessary to restore pictures of full resolution. The functionality of spatial scalability is very important for video transmission through inhomogeneous communication networks, i.e. networks with different transmission bitrates as well as for better protection of video transmission in error-prone environments.

The functionality of spatial scalability is already provided in the MPEG-2 [1,2] and MPEG-4 [3] video compression standards. Unfortunately, the existing solutions standardized by both video coding standards are inefficient because of unacceptably high bitrate overheads as compared to single-layer encoding of video.

There were many attempts to improve spatially scalable coding of video. Among various proposals, application of subband decomposition should be considered as very promising [4-8]. The idea is to split each image into four spatial subbands. The subband LL of lowest frequencies constitutes a base layer while the other three subbands are jointly transmitted in an enhancement layer. Unfortunately, in most of such coders, it is difficult to allocate appropriate number of bits to the base layer and to the enhancement layer. A practical requirement is that the bitstream of the base layer does not exceed the bitstream of the enhancement layer. In order to meet this requirement, a combination of spatial scalability and SNR scalability has been proposed [9].

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

Another approach has been proposed by the authors who introduced a concept of spatio-temporal scalability being a mixture of spatial and temporal scalability [10,11]. This approach leads to two-layer systems where the base layer bitstream constitutes about 40% of the overall bitstream.

Temporal resolution reduction is achieved by partitioning of the stream of B-frames: each second frame is included into the enhancement layer only. Therefore we have two types of B-frames: BE-frames which exist in the enhancement layer only and BR-frames which exist both in the base and enhancement layers.

Table 1. Frames of different types in both layers.

Base	I		BR		P		BR		P
Enhancement	I	BE	BR	BE	P	BE	BR	BE	P

Typical GOP structure is as follows:
I–BE–BR–BE–P–BE–BR–BE– P–BE–BR–BE– P–BE–BR–BE
or
I–BE–BR–BE– BR–BE–P–BE–BR–BE– BR–BE– P–BE–BR–BE. .

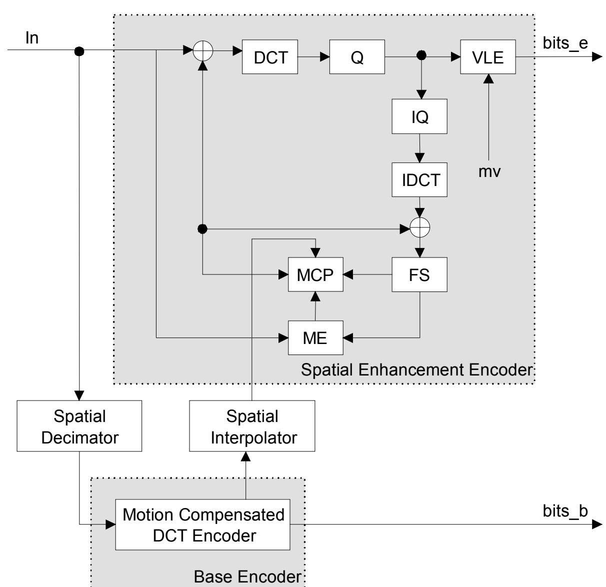
3. CODER STRUCTURE

The I-frames are encoded using subband decomposition. The LL subband (the lowest subband) is transmitted in the base layer while the other three subbands correspond to the enhancement layer.

In multilayer systems, the individual layers of I-frames correspond to group of subands obtained via wavelet decomposition.

The P- and B-frames are encoded using a technique which can be described as a modification of the technique already defined in MPEG-2 standard for spatial scalability (Fig. 1). For each macroblock, a choice between possible temporal predictions and spatial interpolation from the base layer is made. In particular, for BR-frames, the enhancement layer macroblocks are predicted from both reference (past and future) reference frames from the enhancement layer as well from the interpolated frame of the base layer.

Motion estimation is made independently for both layers. Nevertheless, the number of motion vectors estimated in the base layer is only one fourth of that in the enhancement layer.



DCT	Discrete cosine transform
Q	Quantization
IQ	Dequantization
IDCT	Inverse cosine transform
FS	Frame memory
MCP	Motion-compensated predictor
ME	Motion estimator

Figure 1. The overall coder structure.

4. EXPERIMENTAL RESULTS

In order to verify the structure proposed a verification model have been prepared. The software written in C++ language is currently available for progressive sequences with the input resolution defined by the standard digital television resolution. The software runs on Sun 20 workstations under the Solaris operational system as well as on PC-compatible workstations under Windows NT operating system.

The experiments have been made with progressive 720 × 576, 50 Hz, 4:2:0 test sequences (cf. Table 2). A frame from the test sequence is shown in Fig. 2.

Figure 2. A frame from the test sequence *Basketball*

A typical it allocation distribution is shown in Fig. 3. The plot is obtained for consecutive frame in a test video sequence.

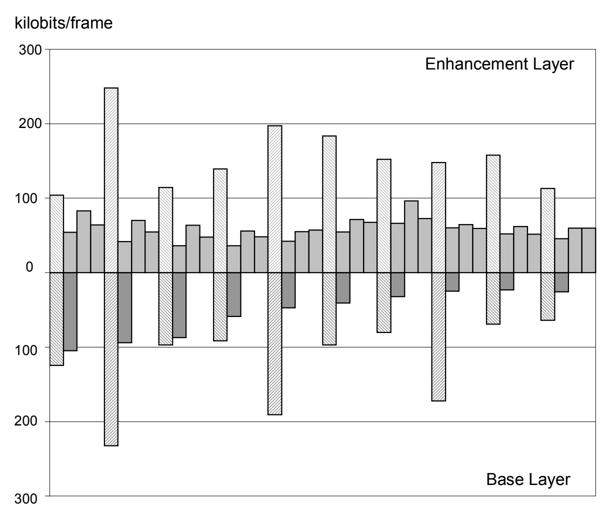


Figure 3. Bits allocated to individual frames rate for proposed coding method: bits allocated to intra-pictures are denoted by slashes, those allocated to P-frames are denoted by backslashes while those corresponding to B-pictures are denoted by solid bars.

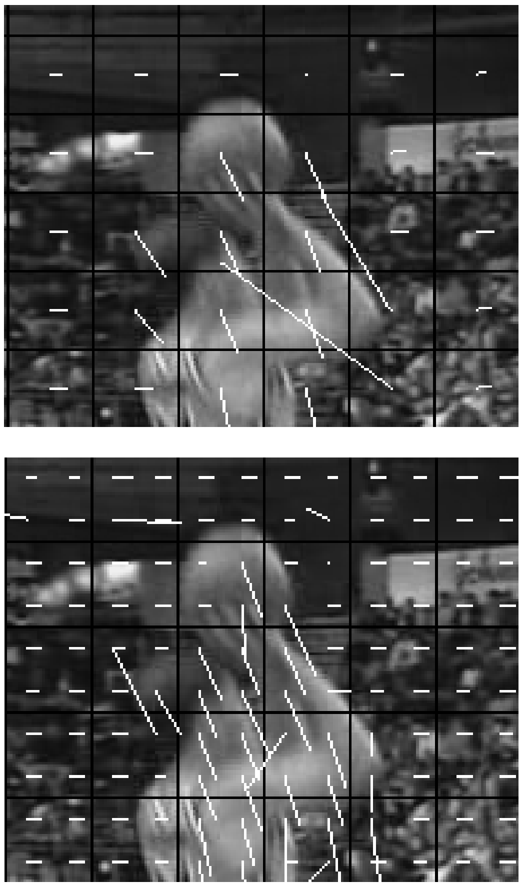


Figure 4. The top picture presents base layer motion vectors, the bottom picture shows enhancement layer motion vectors.

Table 2. The experimental results for the test sequence *Basketball*

Single layer coder (MPEG-2)	Bitsream [Mb]	5,17
	Average PSNR [dB] for luminance	31,36
Proposed scalable coder	Average PSNR [dB] for luminance	31,35
	Base layer bitstream [Mb]	2,14
	Base layer bitstream as percent of the total bitsream	36,2
	Scalability overhead [%]	14,3

Figure 3 as well as the data from Table 2 prove that the base layer bitstream is about 40% total bitstream. This partition is well balanced over all frame types.

Motion vectors in both layers are highly correlated (Figs. 4 and 5). This property can be exploited for further data reduction.

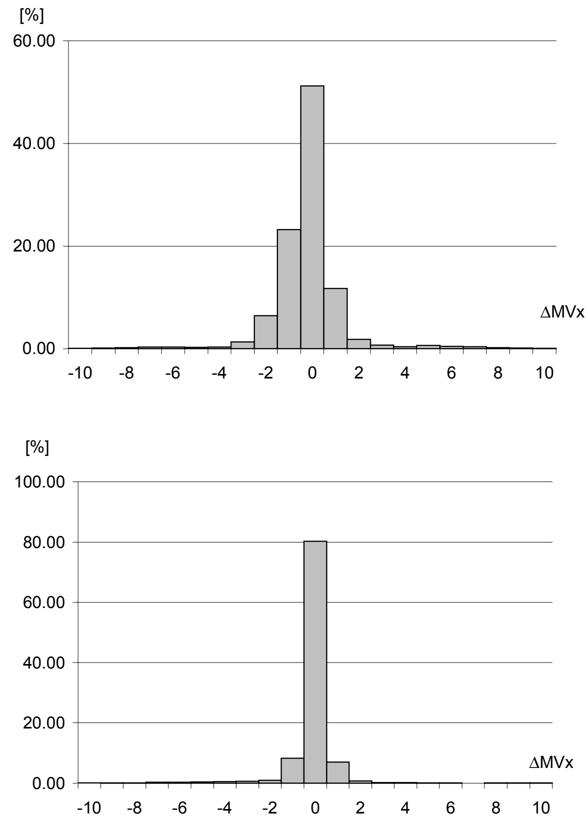


Figure 5. Histograms of rounded differences between the base and enhancement layer motion vectors.
The two plots correspond to two vector components.

5. CONCLUSIONS

The experimental results from Table 2 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 bitrate overhead due to scalability is about 10%. The codec proposed outperforms spatially scalable MPEG-2 [1] or MPEG-4 [3] coders which generate bitrate overheads often exceeding 50%.

The system proposed can be easily extended onto multi-layer systems where the bitrate ratio of consecutive layers is about 1:2.5.

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,” New York: Chapman & Hall, 1997.

[3] International Organization For Standardization, “MPEG-4 Video Verification Model,” ISO/IEC JTC1/SC29/WG11, Stockholm, 1998.

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

[5] F. Bosveld, ”Hierarchical video compression using SBC,” Ph.D. dissertation, Delft University of Technology, Delft 1996.

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

[7] Senbel, H. Abdel-Wahab, ”Scalable and robust image compression using quadtrees,” *Signal Processing: Image Communication*, vol. 14, pp. 425-441, 1999.

[8] Shen, E. Delp, ”Wavelet based rate scalable video compression,” *IEEE Trans. Circuits Syst. Video Technology*, vol. 9, pp. 109-122, February 1999.

[9] U. Benzler, ”Scalable multi-resolution video coding using a combined subband-DCT approach”, *Picture Coding Symposium 1999*, pp. 17-20.

[10] M. Domański, A. Łuczak, S. Maćkowiak, R. Świerczyński, ”Hybrid coding of video with spatio-temporal scalability using subband decomposition,” in *Signal Processing IX: Theories and Applications*, pp. 53-56, Rhodes, September 1998.

[11] M. Domański, A. Łuczak, S. Maćkowiak, R. Świerczyński, Ulrich Benzler „Spatio-Temporal Scalable Video Codecs with MPEG-Compatible Base Layer”, *Picture Coding Symposium 1999*, pp. 45-48.