EFFICIENT STRUCTURE OF VIDEO CODERS
WITH MOTION-COMPENSATED FINE-GRANULARITY SCALABILITY

Marek Domański, Sławomir Maćkowiak, Łukasz Błaszak

Poznań University of Technology, Institute of Electronics and Telecommunication, Poznań, Poland
E-mail: { domanski, smack, lblaszak } @ et.put.poznan.pl

ABSTRACT

The paper describes a two-loop coder structure suitable for mixed spatial and temporal scalability combined with fine-granularity. The system exploits independent motion estimation and compensation for each loop. The additional bitstream related to the increased number of motion vectors is compensated by more precise motion compensated performed independently at each level of resolution. The measures to limit the drift are also included. Drift is restricted to one loop only. Moreover, within a video sequence drift is periodically set to zero. The structure is suitable for encoding of video sequences within the framework of hybrid video coders like MPEG-2/4, H.263 and H.26L. The experimental results are presented in order to prove small scalability overhead in the bitrate, which has been tested for a wide range of bitrates as well as for different reference coders (H.263, MPEG-2).

1. INTRODUCTION

Scalable video coding means that the transmitted bitstream can to be partitioned into some layers in such a way that an arbitrary number of lower layers are decodable into a video sequence with reduced spatial resolution, temporal resolution or signal-to-noise ratio (SNR).

Efficient and flexible scalable video coding is one of the challenging problems of contemporary research on video coding. Wide and important applications include video streaming through error prone channels, especially wireless networks of new generations [1].

The existing video standards MPEG-2 and MPEG-4 provide spatial, temporal and SNR scalability [2,3]. Nevertheless the classic solutions for two- and three-layer scalability are often not enough efficient and flexible.

Moreover, MPEG-4 provides Fine-Granularity-Scalability (FGS) for precise matching the bitstream to channel capacity [4]. The latter solution exploits coders that consist of two parts: the base layer encoder is a motion-compensated hybrid one that produces the base layer bitstream, and the enhancement layer bitstream is generated by a coder with no inner temporal prediction, i.e. the enhancement data are encoded with respect to the low-quality pictures from the base layer. The lack of motion-compensated temporal prediction in the enhancement layer ensures that no accumulating error occur when only a portion of the enhancement bitstream is received by the decoder. Moreover, the bitrate of the received bitstream part can be seamless controlled in a relatively easy-to-implement system.

Unfortunately, the MPEG-4 FGS coders exhibit significant scalability overhead as compared to respective single-layer (non-scalable) coders.

Current research activities in scalability are related to two major groups of approaches [5]: wavelet-based techniques and improvements of the hybrid transform coders. The difficulty with the first approach is related to the necessity of embedding motion-compensated prediction into wavelet-based systems. Nevertheless, the 3-D-wavelet approach seems to be very promising (e.g. [6]) as it provides inherently scalable systems. The second approach [7-13] mostly exploits multi-loop systems. Here, we are dealing with the latter approach. Some similar approaches are those described in [8,13]. The major differences are in independent motion estimation and compensation in both loops, mixed spatio-temporal scalability as well as in modified and improved coding of B-frames.

2. TEMPORAL AND SPATIAL SCALABILITY

In contrary to separate spatial or temporal scalability, the authors have introduced a concept of spatio-temporal scalability being a mixture of spatial and temporal scalability [14]. This approach leads to two-layer systems where the base-layer bitstream can be significantly less than half of the total bitstream. Spatio-temporal scalability can be provided in various picture sequence structures: with or without B-frames and with or without groups of pictures (see Figs. 1 and 2). Some P-frames exploit two reference frames: the previous full-resolution I- or P-frame and the current frame interpolated from the base layer. Such frames are denoted as PI-frames.

In the case of sequences with B-frames, temporal resolution reduction is achieved by partitioning of the stream of B-frames thus leaving some frames in the enhancement layer only. (Fig. 2). In this case, there exist 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.

3. IMPROVED CODING OF B-FRAMES

Improved prediction is used for the BR-frames, which are the B-frames represented in both layers. Each macroblock in a full-resolution BR-frame can be predicted from the following reference frames (Fig.2):
- previous reference frame (I- or P-frame),
- next reference frame (I- or P-frame),
- current reference frame (BR-frame).

The data from the previous and next reference frames are motion-compensated, and data from the current reference frame are upsampled in the two-dimensional space domain. The best

suited reference frame or average of two or three reference frames is chosen according to the criterion of smallest prediction error. Instead of intra-macroblocks, the macroblocks with prediction referring to the interpolated base-layer image are used.

The improvement on standard MPEG-2 prediction within a single layer consists in another decision strategy. The best prediction/interpolation is chosen from all three possible reference frames: previous, future and interpolated [15].
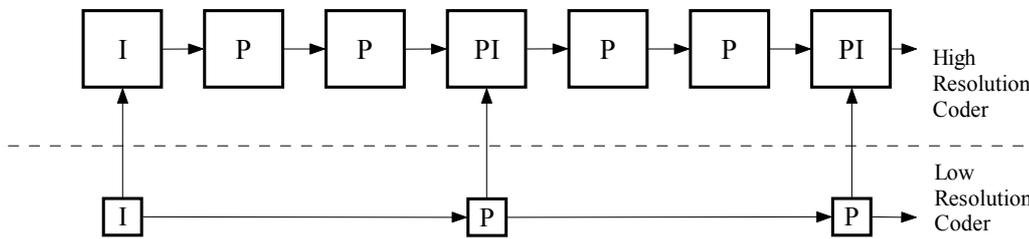

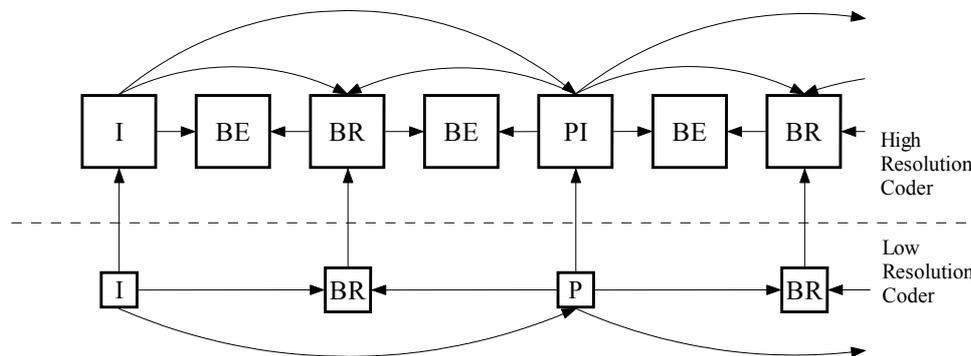Fig 1. Spatio-temporal scalability within a picture sequence with no B-frame and no GOP structure.

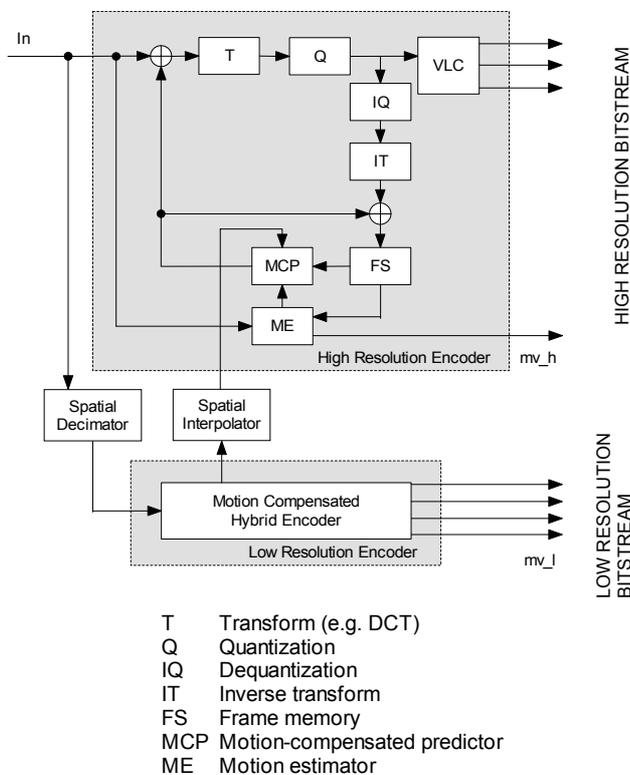Fig. 2. Spatio-temporal scalability within a picture sequence with many B-frames.

T   Transform (e.g. DCT)
Q   Quantization
IQ   Dequantization
IT   Inverse transform
FS   Frame memory
MCP   Motion-compensated predictor
ME   Motion estimator

Fig. 3. The two-loop structure of a coder with spatio-temporal scalability.

4. TWO-LOOP FILTER STRUCTURE

The two bitstreams with two basic resolutions are produced in two independently motion-compensated coder parts (Fig. 3). Usually experimental data prove that the bitrate needed for additional motion vectors is well compensated by the decrease in the number of bits spent for the transform coefficients needed for prediction error encoding [16]. A base-layer macroblock corresponds to four enhancement-layer macroblocks, therefore the respective motion vectors are obviously quit different.

The structure can be extended to three loops in a straightforward way.

5. FINE GRANULARITY AND DRIFT

Fine granularity may be obtained by use of splitting the data produced on any resolution level. For example, motion vectors and the most significant bitplanes may be received while the other bitplanes are lost due to bandwidth decrease. Another option is to transmit first nonzero DCT coefficients from each block (Fig. 4). In that way, the bitstream fed into a decoder may be well matched with the throughput available. It means that the decoding process exploits only a part of one bitstream thus suffering from drift. Always, only one of the bitstreams is split, usually the medium- or high-resolution one. Therefore only one of the bitstreams received is affected by drift.

The phenomenon of drift is related to the reconstruction errors which are accumulating during the process of decoding of the consecutive frames. Therefore insertion of intra-coded frames bounds propagation of drift errors to groups of pictures (GOPs).

In the absence of GOPs (see Fig. 1) some special PIM-pictures can be periodically inserted into the enhancement layer with fine-granularity scalability. In the encoder, this enhancement layer may be additionally reconstructed from the minimum portion of the bitstream thus creating "maximum drifted" version of consecutive pictures. PIM-picture is a picture predicted from this "maximum drifted" layer and from the base layer (by use of spatial interpolation).

Moreover, higher percentage of B-frames also causes that drift accumulates slower.
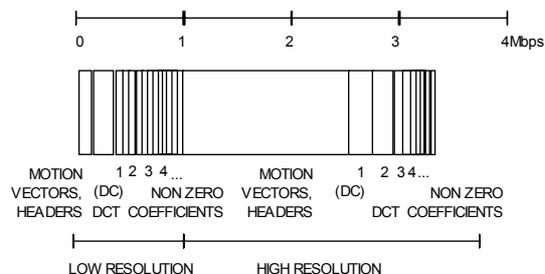
Fig. 4. Bitstream structure for the progressive test sequence *Flower Garden* with total bitrate 3.31 Mbps

6. EXPERIMENTAL RESULTS FOR TWO-LOOP CODERS

The performance of the two loop structure has been tested for various bitrates as well as for two different reference coders: H.263 baseline (Figs.5 and 6) and MPEG-2 MP @ ML (Fig. 7). Progressive sequences have been used for all tests.

For some test sequences and some bitrates chosen, the astonishing feature of the results is that the performance of the two-loop coder i.e. scalable coder, is better than that of the reference single-layer coder. Such results have been obtained independently for both series of experiments based on two different coders and two different sequence structures. The explanation is probably related to the specific sequence structure where the low-resolution bitstream is used as additional reference to each second frame. In some cases, almost 50% macroblocks have exploited interpolation in the prediction process.

Another tests proved convenient allocation of bitrates between both resolution layers (Figs. 6 and 7).
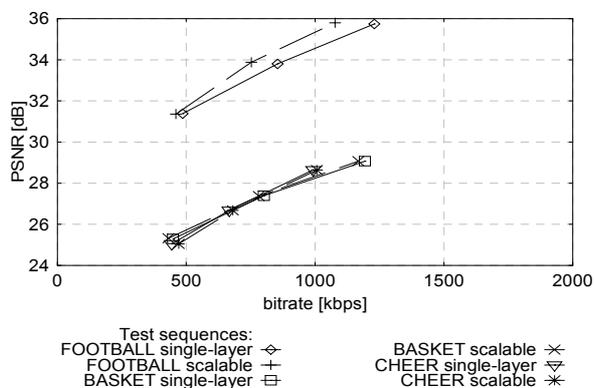
Fig. 5. Coding efficiency (luminance PSNR versus bitrate) for the two-loop structures based on H.263 baseline coder (temporal subsampling factor = 2, image resolution – CIF, 30 fps).
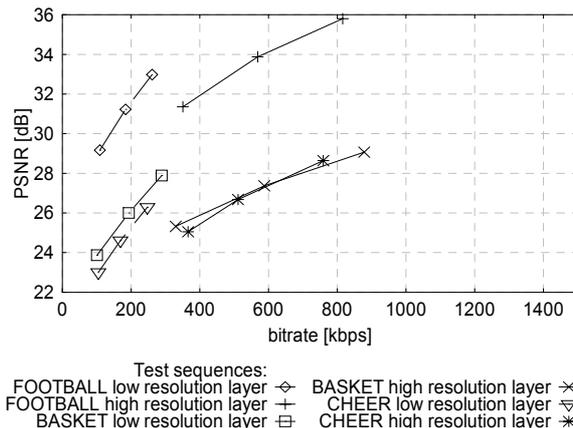
Fig 6. The low-resolution and the high-resolution layers in the two-loop coder based on an H.263 baseline coder (CIF, 30fps).
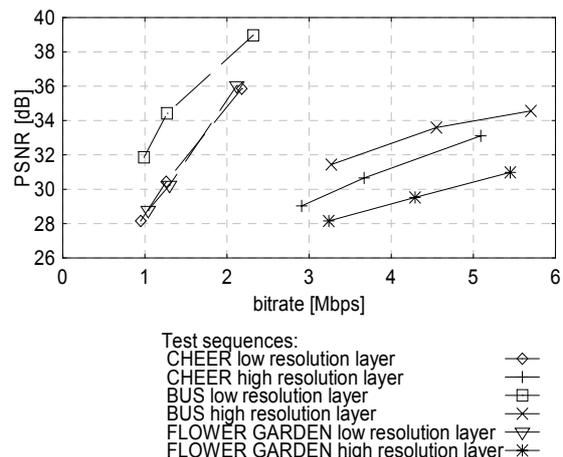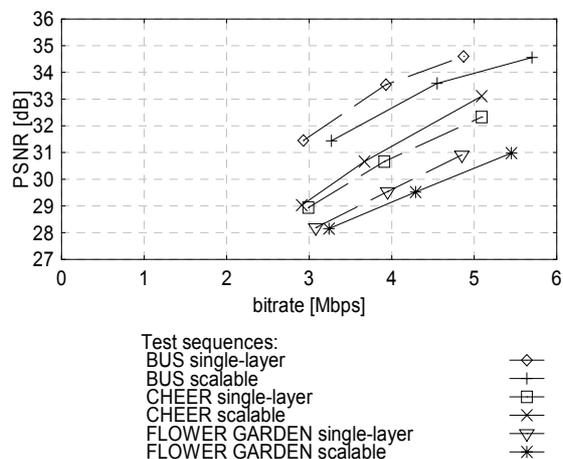
Fig. 7. Coding efficiency (luminance PSNR versus bitrate) for the two-loop structures (temporal subsampling factor = 2, image format = 4CIF, 50 fps). Scalable coders compared to the respective MPEG-2 single-layer coders.

7. EXPERIMENTAL RESULTS FOR FINE GRANURITY AND DRIFT

Besides the overall performance of the scalable coder, the performance for intermediate bitrates has been also measured in order to estimate the efficiency of fine-granularity scalability. For sake of simplicity, the nonzero coefficient allocation scheme for FGS has been implemented only. Application of bitplane coding [4] would improve efficiency of the scalable coder. The number of nonzero DCT coefficients allocated to a given layer controls smoothly the bitrate of the corresponding layer (Fig. 8,9). The respective plots are quite similar for various test sequences.

Drift accumulation is very fast in the absence of B-frames (Fig. 9). Application of BE/BR-frame coding scheme (Fig. 2) reduces significantly the process of drift accumulation.
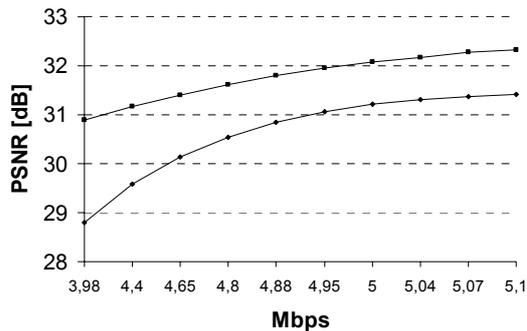

Fig. 8. Compression efficiency of the fine-granularity-scalability implemented in a two-loop coder (lower curve) compared to that of single layer MPEG-2 (upper curve). Results obtained for the test sequence *Funfair* with total bitrate 5 Mbps and the base layer bitrate about 1.66 Mbps.
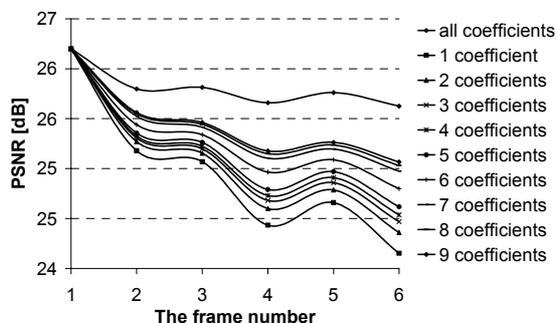

Fig. 9. Decreasing signal-to-noise ratio according to drift for various numbers of DCT coefficients per block transmitted in the enhancement layer to the decoder. Test sequence *Funfair* with an average bitrate 0.5 Mbps and without B-frames.

8. CONCLUSIONS

Described is a two-loop scalable coder. The major differences with respect to other proposals [8,10,12] are: mixed spatio-temporal scalability, independent motion estimation for each motion-compensation loop, i.e. for each spatio-temporal resolution layer, BR/BE-frame structure and improved prediction of BR-frames. These features are also the reasons for very good performance of the whole coder.

The encoded bitstream syntax is almost standard MPEG or H.263 one. The bitrate of the base layer can be smoothly controlled starting from below 15% of the total bitrate. With the same bitrate as by MPEG-2 main (nonscalable) profile or H.263 baseline coder, the scalable coder proposed reaches almost the same quality. The bitrate overhead due to scalability is mostly below 10% and quit often is even negative.

REFERENCES

[1] D. Wu, Y. Hou, Y. Zhang, "Scalable video coding and transport over broad-band wireless networks," *Proc. of the IEEE*, vol. 89, pp. 6-20, January 2001.

[2] ISO/IEC IS 13818, "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information".

[3] ISO/IEC IS 14496, "Information Technology - Generic Coding of Audiovisual Objects".

[4] ISO/IEC JTC1/SC29/WG11 N3908, "MPEG-4 Video Verification Model, ver. 18.0," Pisa, January 2001.

[5] J.-R.Ohm, M. Beermann, „Status of scalable technology in video coding", Doc. ISO/IEC JTC1/SC29/WG11 MPEG01/M7483, Sydney, July 2001

[6] V. Bottreau, M. Benetiere, B. Felts, and B. Pesquet-Popescu, A fully scalable 3D subband video codec, *IEEE Conference on Image Processing 2001, ICIP'01*, vol. II, pp. 1017-1020,Thessaloniki, October 2001.

[7] J.-R.Ohm, M. van der Schaar, *Scalable Video Coding*, Tutorial material, *Int. Conf. Image Processing ICIP 2001*, IEEE, Thessaloniki, October 2001.

[8] Y. He, R. Yan, F. Wu, S. Li, "H.26L-based fine granularity scalable video coding," Doc. Doc. ISO/IEC JTC1/SC29/WG11 MPEG01/m7788, December 2001.

[9] K.Rose, S. Regunathan, "Toward optimality in scalable predictive coding," *IEEE Trans. Image Proc.*, vol. 10, pp.965-976, July 2001.

[10] M. Domański, S. Maćkowiak, „Modified MPEG-2 video coders with efficient multi-layer scalability," *Int. Conf. Image Processing ICIP 2001*, IEEE, vol. II, pp. 1033-36, Thessaloniki, October 2001.

[11] M. Domański, A. Łuczak, S. Maćkowiak, „Spatio-temporal scalability for MPEG video coding", *IEEE Trans. Circ. and Syst. Video Technology*, vol. 10, pp. 1088-1093, Oct. 2000.

[12] A. Reibman, L. Bottou, A. Basso, "DCT-based scalable video coding with drift," *Int. Conf. Image Processing ICIP 2001*, IEEE, vol. II, pp. 989-992, Thessaloniki, October 2001.

[13] U. Benzler, "Spatial scalable video coding using a combined subband-DCT approach", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, pp. 1080-1087, October 2000.

[14] 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, Proc. EUSIPCO-98*, pp. 53-56, Rhodes, September 1998.

[15] M. Domański, A. Łuczak, S. Maćkowiak, "Scalable MPEG Video Coding with improved B-frame prediction", *Proceedings of IEEE International Symposium Circuits and Systems ISCAS 2000*, vol. II, pp. 273-276, Geneva, May 2000.

[16] A. Łuczak, S. Maćkowiak, M. Domański, "Spatio-temporal scalability using modified MPEG-2 predictive video coding", *Signal Processing X: Theories and Applications, Proc. EUSIPCO-2000*, pp. 961-964, Tampere, September. 2000.