FINE GRANULARITY IN MULTI-LOOP HYBRID CODERS WITH MULTI-LAYER 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 generic multi-loop coder structure suitable for mixed spatial and temporal scalability combined with fine granular SNR scalability. The structure is suitable for various variants of hybrid video coders like MPEG-2, H.263 and H.26L. The idea of mixed spatial and temporal scalability i.e. spatio-temporal scalability is substantial for the proposal. Its application allows improving the scalable coding efficiency i.e. decreasing the scalability overhead. The coder consists of independently motion-compensated sub-coders that produce bitstreams corresponding to individual levels of spatio-temporal resolution. The bitrate can be smoothly matched to the particular channel bandwidth by use of data partitioning, which is related to drift errors in the decoder. Accumulation and propagation of these errors can be bounded by use of proper structure of groups of pictures.

1. INTRODUCTION

Emerging new audiovisual services are related to real-time streaming of video that has requirements on quality of service, which vary between sections of heterogeneous communication networks. Therefore the multicast video transmission needs scalability in order to match the video transmission bitrate with channel throughput available. The need for scalable video transmission becomes crucial for wireless networks with their unreliability and bandwidth fluctuations [1,2].

Recently, the MPEG-4 [3] has adopted Fine-Granularity-Scalability (FGS) as a tool for precise tuning a layer bitstream to channel payload. In the MPEG-4 FGS, the intra-frame coding of the enhancement layer is very flexible but not too efficient. Therefore some solutions have been already proposed [4] that exploit interframe coding and motion compensation for the enhancement layer with FGS [5-10] while the others are exploiting the wavelet approach [4].

Here, we are dealing with hybrid motion-compensated coder structures with fine-granularity-scalability such that the coding efficiency is high and possibly close to that of single-layer coders. The approach exploits multi-loop structures and limitation of drift propagation to a number of frames.

The work has been partially supported by the IST Programme of the EU under contract number IST-2000-26467 (MASCOT).

The generic structure is an extension of that proposed in [9] and differs from that described in [7] by mixed spatio-temporal scalability and independent motion estimation and compensation performed in the individual prediction loops. These two features together with other improvements described further are substantial for high efficiency obtained.

2. MIXED SPATIAL AND-TEMPORAL SCALABILITY WITH FINE GRANULARITY IN HYBRID CODERS

The scalable coder consists of two or three motion-compensated coders (Fig. 1) that encode a video sequence and produce two or three bitstreams corresponding to two or three different levels of spatial and temporal resolution (Fig. 1).

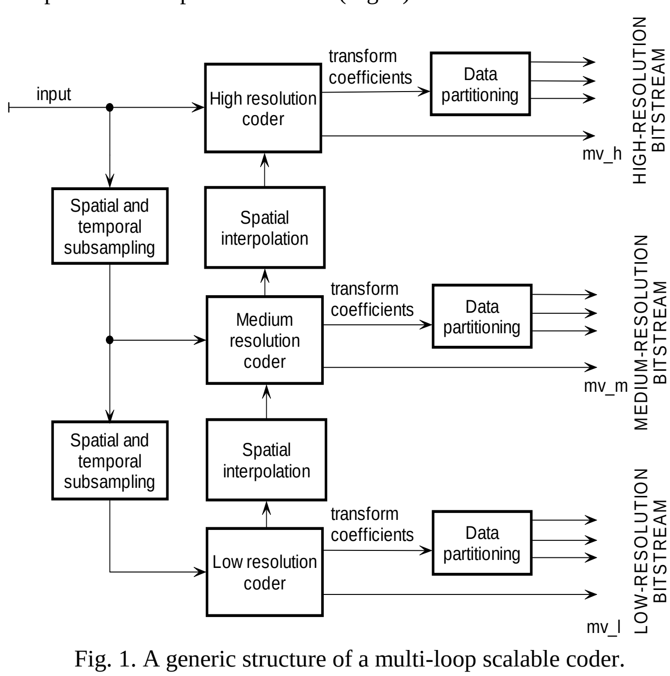

Fig. 1. A generic structure of a multi-loop scalable coder.

Each of the coders has its own prediction loop with own motion estimation. Such a structure may seem to be redundant with respect to the number of motion vectors estimated and transmitted. Nevertheless previous experiments have proved that the optimum motion vectors are different at different spatio-temporal resolutions [11]. Usually experimental data prove that the bitrate needed for additional motion vectors is well compensated by the decease in the number of bits spent for the transform coefficients needed for prediction error encoding [11,12]. Moreover, the overhead related to motion vectors is only less than 25% for two-loop code and less than 31% for a three-loop coder.

Fine granularity may be obtained by use of splitting the data produced on any resolution level. 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). Moreover, higher percentage of B-frames also decreases the influence of drift.

There are various sequence structures possible. In a not favorable case of absence of GOPs and B-frames, the enhancement layer sequence can be divided into GOPs in order to limit the process of drift propagation (Fig. 2).

In order to improve coding efficiency the prediction scheme from Fig. 3 has been proposed [13,14]. Here, the frames skipped frames are only B-frames (called BE-frames). Those B-frames that exist in both layers (BR-frames) are also used as reference frames. Moreover an improved prediction scheme [14] can be used.

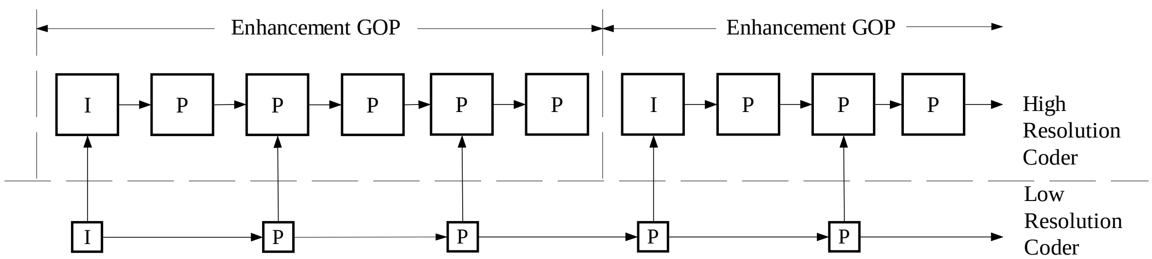

Fig. 2. Exemplary structure of a video sequence: No B-frames, just one I frame in the base layer and GOPs in the enhancement layer.

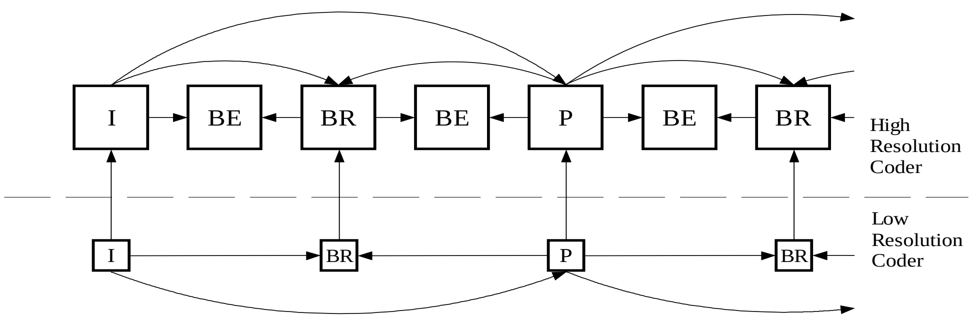

Fig. 3. Exemplary structure of a video sequence: Number of B-frames is 75% of the total number of frames; both the base and the enhancement layer divided into GOPs.

3. EXPERIMENTAL RESULTS

The performance of the two loop structure has been tested for various bitrates. Progressive sequences have been used for all tests.

Two basic series of experiments have been performed:

a) H.263-based experiments have made with CIF sequences of the structure from Fig. 2 but without GOPs in the enhancement layer. The coder used was built on the H.263 baseline coder.

b) MPEG-2-based experiments used 4CIF sequences with the structure from Fig. 3 and the GOP length of 12 for both layers.

The overall coding performance is summarized in Table 1. The figures are average PSNR values as well as average bitrates for selected test sequences. The values for two-layer bitstreams have been compared to single-layer bitstreams obtained using standard MPEG-2 or H.263 coders with the same options switched on. The values at the output of low-resolution coder are also included in Table 1.

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.

Fine granularity has been obtained by transmitting only a desired portion of the DCT-data from the bitstream of the highest resolution. This can be efficiently done on the basis of bit-planes.

Table 1. The two-loop coder performance measured for whole resolution levels:
Results obtained for progressive sequences. The H.263 coder without PB-frames, respective scalable coder without GOPs in both layers. The MPEG-2 coder and respective scalable coder with GOP length of 12 frames.

H.263 - based coder for CIF (352 × 288) sequences		*Football*	*Basket*	*Cheer*	*Fun*	*Bus*
Single-layer coder (H.263)	Bitstream [Kbps]	428.64	422.14	402.33	424.86	414.61
	Average luminance PSNR [dB]	29.58	26.04	25.62	26.30	27.57
Proposed scalable coder	Low resolution layer bitstream [Kbps]	99.27	103.17	98.28	103.89	99.70
	Low resolution layer average PSNR [dB] for luminance	26.76	24.64	23.19	24.17	26.06
	High resolution layer bitstream [Kbps]	319.20	330.67	315.41	313.13	302.61
	Average PSNR [dB] for luminance recovered from both layers	29.55	26.02	25.58	26.29	27.59
Single-layer coder (H.263)	Bitstream [Kbps]	875.77	775.02	692.31	831.36	790.55
	Average luminance PSNR [dB]	32.38	27.82	27.82	28.85	29.69
Proposed scalable coder	Low resolution layer bitstream [Kbps]	190.22	194.22	179.08	197.55	190.99
	Low resolution layer average PSNR [dB] for luminance	29.32	26.68	25.31	26.61	28.37
	High resolution layer bitstream [Kbps]	578.17	617.17	546.91	594.46	567.83
	Average PSNR [dB] for luminance recovered from both layers	32.42	27.83	27.83	28.85	29.68
Single-layer coder (H.263)	Bitstream [Kbps]	1271.28	1139.21	998.14	1216.35	1121.57
	Average luminance PSNR [dB]	34.24	29.28	29.55	30.62	31.14
Proposed scalable coder	Low resolution layer bitstream [Kbps]	282.83	287.39	260.14	292.54	283.68
	Low resolution layer average PSNR [dB] for luminance	31.21	28.42	27.06	28.51	30.25
	High resolution layer bitstream [Kbps]	851.84	902.34	788.41	877.47	835.02
	Average PSNR [dB] for luminance recovered from both layers	34.27	29.28	29.59	30.62	31.15
MPEG-2 - based coder for 4CIF (704 × 576) sequences		*Cheer*	*Flower Garden*	*FunFair*	*Stefan*	*Bus*
Single-layer coder (MPEG-2)	Bitstream [Mbps]	2.99	3.08	3.17	2.96	2.93
	Average luminance PSNR [dB]	28.94	28.19	29.18	31.99	31.45
Proposed scalable coder	Low resolution layer bitstream [Mbps]	0.95	1.04	0.77	0.98	0.99
	Low resolution layer average PSNR [dB] for luminance	28.15	28.78	27.67	32.88	31.86
	Total bitstream [Mbps]	2.91	3.24	3.23	2.99	3.27
	Average PSNR [dB] for luminance recovered from both layers	29.03	28.15	29.21	32.06	31.44
Single-layer coder (MPEG-2)	Bitstream [Mbps]	3.91	3.95	3.91	3.89	3.93
	Average luminance PSNR [dB]	30.66	29.54	30.84	33.84	33.54
Proposed scalable coder	Low resolution layer bitstream [Mbps]	1.26	1.30	1.50	1.27	1.27
	Low resolution layer average PSNR [dB] for luminance	30.43	30.24	32.77	35.73	34.42
	Total bitstream [Mbps]	3.67	4.29	3.80	3.71	4.55
	Average PSNR [dB] for luminance recovered from both layers	30.66	29.52	30.79	33.74	33.59
Single-layer coder (MPEG-2)	Bitstream [Mbps]	5.09	4.85	4.76	4.74	4.87
	Average luminance PSNR [dB]	32.33	30.91	32.17	35.20	34.60
Proposed scalable coder	Low resolution layer bitstream [Mbps]	2.18	2.12	1.66	1.93	2.32
	Low resolution layer average PSNR [dB] for luminance	35.86	36.04	33.75	40.51	38.96
	Total bitstream [Mbps]	5.09	5.45	5.09	5.02	5.7
	Average PSNR [dB] for luminance recovered from both layers	33.11	30.98	32.15	35.14	34.56

Here, in order to have a simple implementation, the partitioning has been simply done by limiting the maximum number of the DCT coefficients transmitted per block.

The comparison for intermediate bitrates proves that also fine-granularity scalability is related to acceptable performance (Fig. 4). In the sequences with B-frames, drift remains bounded within a GOP (Fig. 5).

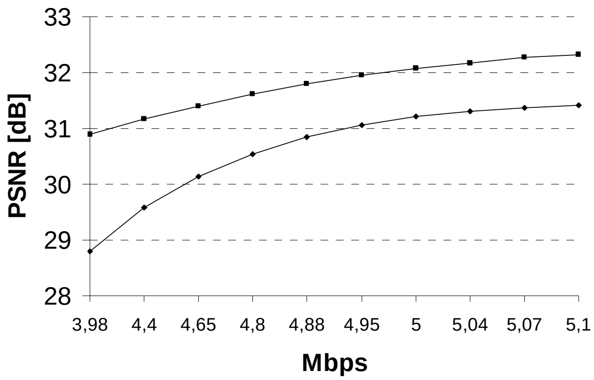

Fig. 4. Compression efficiency of the fine-granularity-scalability implemented in a two-loop coder (lower curve) compared to that of single layer MPEG-2 (upper curve). Results obtained for the test sequence *Funfair* with total bitrate 5 Mbps and the base layer bitrate about 1.66 Mbps.

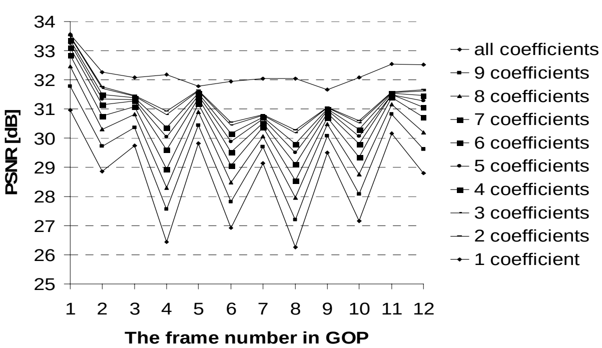

Fig. 5. Decreasing signal-to-noise ratio according to drift for various numbesr of DCT coefficients per block transmitted in the enhancement layer to the decoder. The two-loop coder and the sequence structure from Fig. 3. Test sequence *Funfair* with an average bitrate 5Mbps.

4. CONCLUSIONS

Described is a generic multi-loop coder structure for motion-compensated fine-granularity scalability. The major differences with respect to the proposal from [7] 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,
- improved prediction of BR-frames [14].

These features are also the reasons for very good performance of the whole coder.

For lower bitrates, the lost of quality due to fine granularity is not dramatic. In many applications, the reduced bitrate corresponds to some perturbance and therefore certain loss of coding performance can be acceptable when the whole coder performs well.

5. REFERENCES

[1] L. Hanzo, P. Cherriman, J. Streit, *Wireless Video Communications*, IEEE Press, New York, 2001..

[2] 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.

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

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

[5] S. Li, F. Wu, Y. Zhang, “Experimental results with progressive fine granularity scalable (PFGS) coding,” Doc. ISO/IEC JTC1/SC29/WG11 m5742, March 2000.

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

[7] 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.

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

[9] 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.

[10] 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.

[11] M. Domański, A. Łuczak, S. Maćkowiak, “On improving MPEG spatial scalability,” *Proc. Int. Conf. Image Proc. ICIP 2000*, IEEE, vol. 2, pp. 848-851, Vancouver, September 2000.

[12] 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.

[13] 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, October 2000.

[14] 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.