EFFICIENT HYBRID VIDEO CODERS WITH SPATIAL AND TEMPORAL 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 deals with an efficient coder structure being appropriate for scalable coding of video. The coder consists of two motion-compensated hybrid coders with independent motion estimation and compensation. The structure implements spatial scalability or mixed spatial and temporal scalability that can be combined with fine granular SNR scalability. The encoder exhibits extended capabilities of adaptation to network throughput. The MPEG-2 and H.263 video coding standards are used as a reference but the results are also applicable to the MPEG-4 and H.26L systems with minor modifications. The coder exhibits high level of compatibility with standard H.263 and MPEG 2/4 coders.

1. INTRODUCTION

The well-known classic video coding schemes have been developed and standardized mostly for wireline quasi-error-free transmission systems. Recently, the emergence of broadband wireless networks and related new multimedia services yield new requirements for video coding systems that have to be adapted to unreliable wireless systems with their fades and transmission errors. Moreover, wireless transmission systems exhibit bandwidth fluctuations due to time and receiver position. Typical video coders of H.263 [1] and MPEG-2/4 [2,3] are not enough efficient for video streaming [4,5] in such error-prone environments unless they exhibit the functionality of scalability.

Scalability means that a video data bitstream is partitioned into layers in such a way that the base layer is independently decodable into a video sequence with reduced spatial resolution, temporal resolution or signal-to-noise ratio (SNR). Enhancement layers provide additional data necessary for video reproduction with higher spatial resolution, temporal resolution or signal-to-noise ratio. This functionality is called spatial, temporal or SNR scalability, respectively, as defined by video coding standards: MPEG-2 [2] and MPEG-4 [3]. In the case of bandwidth decrease, the receiver decodes only the base part of the bitstream.

Unfortunately, the scalable coding schemes provided by MPEG-2 and MPEG-4 are not satisfactory in some aspects, like coding efficiency and bandwidth adaptation flexibility. Although MPEG-4 [3] has adopted Fine-Granularity-Scalability (FGS) as a tool for precise tuning a bitstream to channel payload, its coding efficiency is not satisfactory because of lack of temporal prediction in the enhancement layer.

There were many attempts to improve spatially scalable coding of video. Great expectations are related to the inherently scalable wavelet-based techniques [6,7], which have been successfully exploited for flexibly scalable still image compression in the new international standard JPEG 2000 [8]. Unfortunately, in video coding, motion-compensated wavelet-based schemes are still not so successful. Another group of techniques exploits the hybrid coder structures based on motion-compensated prediction and transform block coding [7,9-11,16,17]. A similar approach has been proposed by the authors who introduced a concept of spatio-temporal scalability being a mixture of spatial and temporal scalability [12-14]. This approach was quit successful but mixing this technique with FGS provides even more flexible structure of the encoder [15].

The paper deals with an efficient coder structure that consists of two motion-compensated hybrid coders with independent motion estimation and compensation. The structure implements spatial scalability or mixed spatial and temporal scalability that can be combined with fine granular SNR scalability. The encoder exhibits extended capabilities of adaptation to network throughput. The MPEG-2 and H.263 video coding standards are used as a reference but the results are also applicable to the MPEG-4 and H.26L systems with minor modifications. The coder exhibits high level of compatibility with standard H.263 and MPEG 2/4 coders.

2. CODER STRUCTURE

The scalable coder consists of two motion-compensated coders (Fig. 1) that encode a video sequence and produce two bitstreams corresponding to two different levels of spatial and temporal resolution (Fig. 1). 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 (Fig. 2). Nevertheless previous experiments have proved that the optimum motion vectors are different at different spatio-temporal resolutions. The experimental data prove that usually 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 [13,14].

Most of the hybrid two-loop scalable coders described in the references use the same motion vectors in both motion-compensated prediction loops [7,9-11,16,17]. The application of

independent motion estimation and compensation is characteristic for the proposal considered.

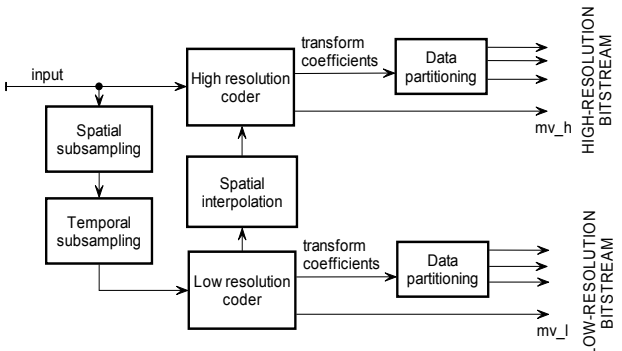
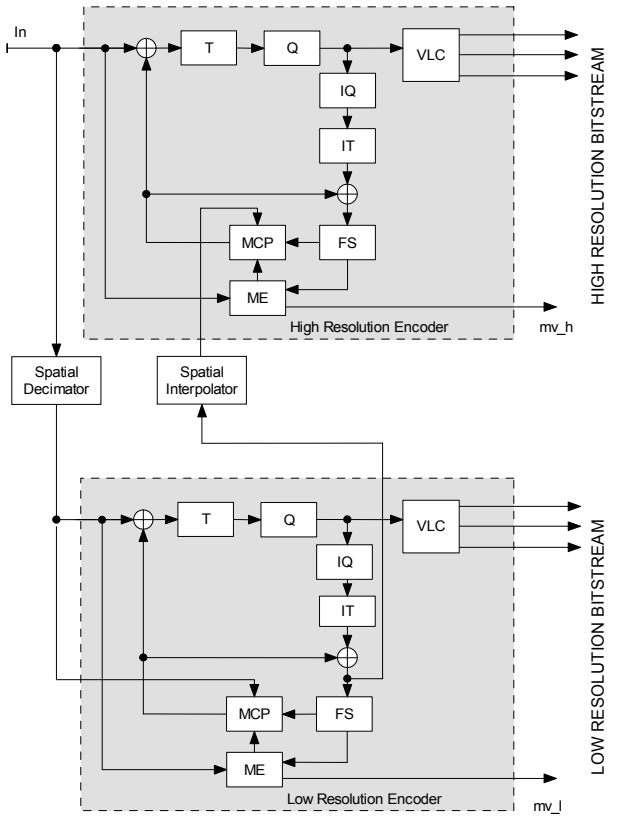


Fig. 1. The general structure of the coder proposed. *m_l, mv_h* - motion vectors.



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

Fig. 2. Detailed scheme of the encoder (temporal subsampling is not included in this figure). *VLC – variable-length coder*.

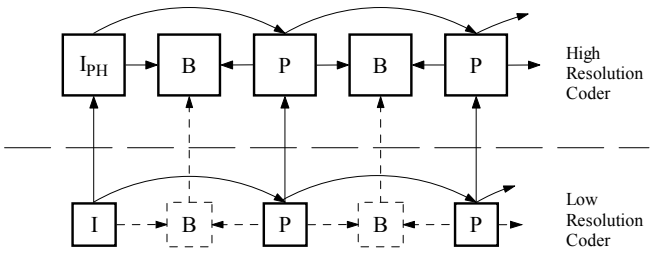


Fig. 3. A picture sequence structure: dotted lines correspond to the absence of temporal scalability. Arrows represent predictions.

3. MODIFIED PREDICTION

The prediction of the images in the enhancement layer exploits both full-resolution images shifted in time as well as the present image interpolated from the base layer (Fig. 3). The coding efficiency is improved by use of a modified prediction scheme [13], which differs from the classic MPEG one by more flexible selection of the reference macroblock.

An enhancement-layer macroblock can be predicted from the following reference frames:

- previous reference frame (for P- and B-frames),
- next reference frame (for B-frames only),
- current reference frame (interpolated).

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.

The improvement on standard MPEG-2 prediction within a single layer consists in another decision strategy. The best prediction/interpolation is chosen from all seven possible reference frames: previous, future, current interpolated, three averaged reference image pairs and an average of the three reference frames [18]. The experimental results prove significant reduction of bitrate being often higher than 10% as compared to the respective with standard MPEG prediction. Especially, the averages of previous and interpolated reference frames are often used in the P-frames. The percentage of such reference macroblocks used often exceeds 40% of the total number of macroblocks.

4. 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. 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 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). It is enough to have GOPs in the enhancement layer only if the high-resolution bitsream is partitioned for fine granularity scalability. It means that such GOPs begin with I-frames in the enhancement layer. Coding of these I-frames exploits reference macroblocks from the interpolated base-layer (low-resolution) images.

In the absence of GOPs some special PI-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. PI-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.

5. 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 with CIF sequences without GOPs in the enhancement layer. The coder used was built on the H.263 baseline coder with no PB-frames.

b) MPEG-2-based experiments used BT.601/4CIF sequences with the structure from Fig. 3 and the GOP length of 12 for both layers.

The coders with both spatial and temporal scalability has been tested. Both temporal and spatial subsampling factor was set to 2. It means that the sequences were:

a) Enhancement layer: progressive CIF 30 Hz,
Base layer: progressive QCIF 15 Hz.

b) Enhancement layer: progressive BT.601/4CIF 50 Hz,
Base layer: progressive CIF 25 Hz.

The overall coding performance is summarized in Figs. 4 and 5. The PSNR values are average values for luminance in 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 (Figs. 6 and 7).

The experimental results prove very good performance of the coder. The bitrate overhead due to scalability is almost always below 10%. For some test sequences and some bitrates chosen (Fig. 6), 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 same phenomenon has been already described and explained in other similar structures [19].

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. This can be efficiently done on the basis of bit-planes.

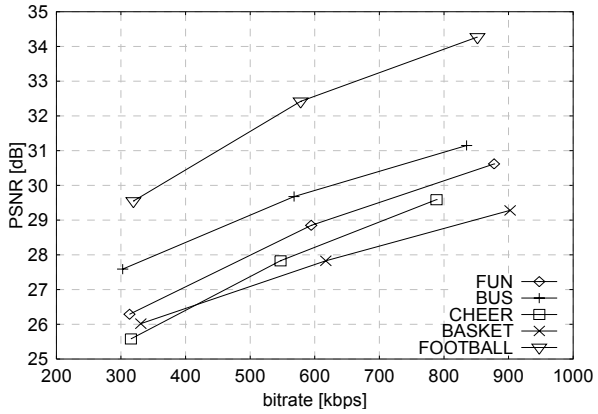


Fig. 4. Performance of a two-loop (two-layer) coder based on the H.263 baseline coder. Plots obtained for progressive 30 Hz CIF sequences.

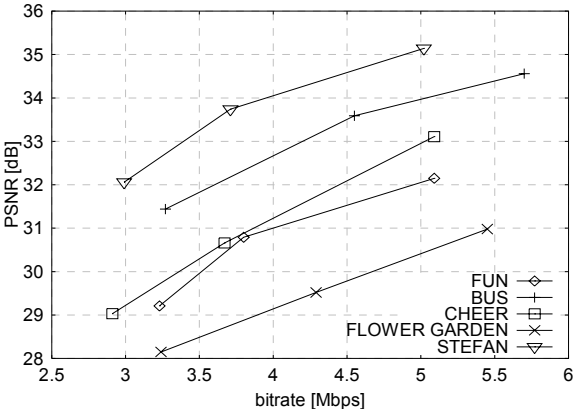


Fig. 5. Performance of a two-loop (two-layer) coder based on the MPEG-2 coder. Plots obtained for progressive 50 Hz BT.601 sequences.

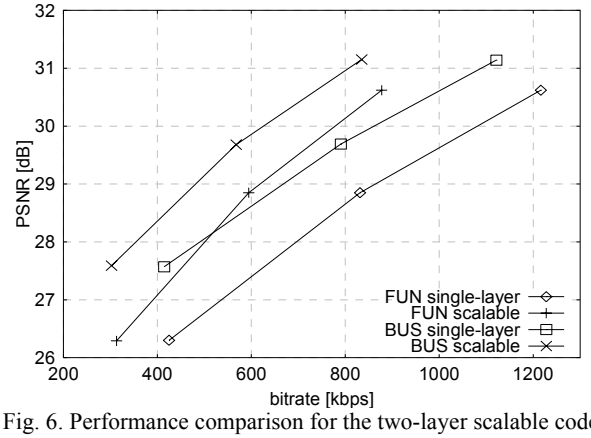


Fig. 6. Performance comparison for the two-layer scalable coder and the H.263 single-layer (nonscalable) baseline coder.

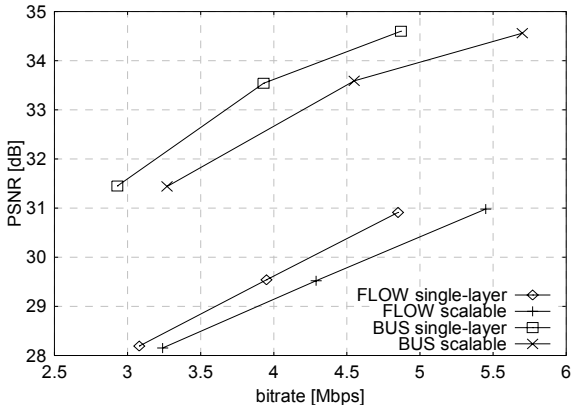


Fig. 7. Performance comparison for the two-layer scalable coder and the MPEg-2 single-layer (nonscalable) coder.

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 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). The respective plots are quite similar for various test sequences.

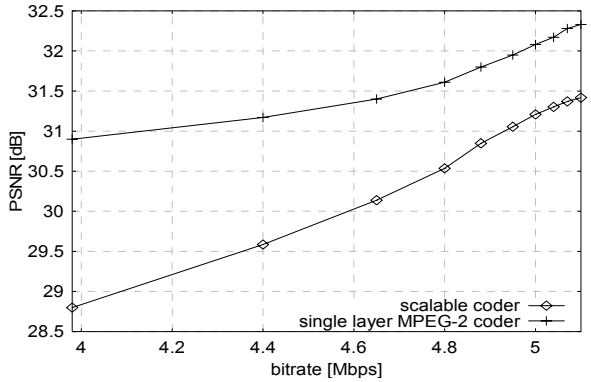


Fig. 8. The fine-granularity-scalability in a two-loop coder (lower curve) compared to a single layer MPEG-2 coder (upper curve). Test sequence *Funfair*, total bitrate 5 Mbps, base layer bitrate about 1.66 Mbps, GOP=12.

6. CONCLUSIONS

Described is a two-layer scalable coder with the functionality of fine granularity. The major differences with respect to other proposals [7,9-11,16,17 and others] are: mixed spatio-temporal scalability, independent motion estimation for each motion-compensation loop and improved prediction of B-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. The bitrate overhead due to scalability is mostly below 10%.

7. REFERENCES

[1] ITU-T, *Video coding for narrow telecommunication channels at < 64kbit/*s, Recommendation H.263, 1996.

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

[3] International Organization For Standardization, "MPEG-4 Video Verification Model v. 18.0", ISO/IEC JTC1/SC29/WG11 N3908, Pisa, January 2001.

[4] 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.

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

[6] P. Topiwala (ed.), *Wavelet image and video compression*, Boston, Kluwer 1998.

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

[8] ISO/IEC IS 15444-1 / ITU-T Rec. T.800, "JPEG 2000 image coding system."

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

[10] 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.

[11] U. Benzler, "Spatial scalable video coding using a combined subband-DCT approach", *IEEE Trans. Circuits Syst. for Video Techn.*, vol. 10, pp. 1080-1087, Oct. 2000

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

[13] M. Domański, A. Łuczak, S. Maćkowiak, "On improving MPEG spatial scalability," *Proc. Int. Conf. Image Proc. ICIP 2000,* IEEE, pp. 2/ 848-851, Vancouver, Sept. 2000.

[14] 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, Sept.. 2000.

[15] 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.

[16] 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.

[17] 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, Oct. 2001.

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

[19] G. Cote, B. Erol, M. Gallant, F. Kossentini, "H.263+: video coding at low bit rates", *IEEE Trans. Circ. and Syst. Video Technology*, vol. 8, pp. 849-865, November 1998.