Multi-loop spatio-temporal scalable video coders

Łukasz Błaszak* Marek Domański* Sławomir Maćkowiak*

Abstract – **This paper presents an original proposal for spatio-temporal scalability in hybrid DCT-based video encoders. The proposed solution has been examined in three scalable encoders based on H.263 standard, MPEG-2 standard and forthcoming AVC video coding standard. The low resolution base layer bitstream is fully compatible with appropriate standard. The whole structure exhibits high level of compatibility with individual building blocks of appropriate encoders and the enhancement bitstream exploits the bitstream semantics and syntax, with some modifications.**

1 INTRODUCTION

The multimedia development forecasts point the networked applications and services in wireless and heterogeneous communication networks. The levels of Quality of Service are different in individual network sections and they even often vary in time. Therefore media adaptation is an important issue for modern multimedia systems. One of the media adaptation issues is video bitsream scalability. Scalability of video means the ability to achieve a video of more than one resolution or quality simultaneously. Scalable video coding involves generating a coded representation (bitstream) in a manner that facilitates the derivation of video of more than one resolution or quality from this bitstream. The scalable video bitsream consists of layers. The base layer corresponds to a video sequence with the lowest quality or resolution. The base layer bitsream may be decoded independently from the other layers of a scalable bitsream. The other layers, the enhancement layers, may be used to achieve higher quality or resolution of the pictures decoded. For a given overall decoded video quality, scalable coding performance is acceptable, if the bitrate is not significantly greater than the bitrate achieved in single-layer coding.

The existing video compression standards define scalable profiles, which exploit classic Discrete Cosine Transform-(DCT)-based schemes with motion compensation. Unfortunately, spatial scalability as proposed by the MPEG-2 coding standard [1,2] is inefficient because the bitrate overhead is too large. Moreover, the solutions of MPEG-4 [13] are also not enough efficient. The new AVC [14] standard defines more efficient coding schemes but still with no scalability.

The paper deals with an original scheme of scalable video coding. The technique is applicable for spatial and temporal scalability. Its applications to MPEG-2, H.263 and AVC video coding systems are considered in the paper. The implementations of scalability in these three coding systems are considered and tested experimentally in order to asses coding efficiency.

2. SPATIO-TEMPORAL DECOMPOSITION

Among various possibilities, the combination of spatial and temporal scalability called spatio-temporal scalability seems very promising [11]. Spatio-temporal decomposition allows to encode the base layer with a smaller number of bits because the base layer corresponds to reduced information.

Spatio-temporal scalability has been proposed in several versions, in particular:

- with 3-D spatio-temporal subband decomposition [8-10],
- with 2-D spatial subband decomposition and partitioning of B-frames data [10,11],
- exploiting as reference frames the interpolated low resolution images from the base layer [12,13].

The choice of the spatial decimator and interpolator has substantial impact on the overall coding efficiency. In the experiments, for decimation, the FIR lowpass zero-phase 7-tap for H.263 and MPEG based scalable encoders and 13-tap filters for AVC scalable encoder have been applied.

The basic structure of a group of pictures (GOP) consists of I- P- and B-frames (Fig. 1). The variant with 3 B-frames between two consecutive I- or P-frames has been chosen because of simple temporal decimation with factor 2.

In the proposed spatio-temporal scalability, the temporal resolution reduction is achieved by partitioning the stream of B-frames: each second frame is skipped in the low resolution encoder.

Therefore there may exist two types of B-frames:

- BE-frames that exist in the enhancement layer only and
- BR-frame that exist both in the base and in the enhancement layer.

The latter may be predicted using interpolation from the decoded low-resolution base-layer frames. In this proposal, improved B-frame encoding [11,12] was used in the enhancement layer, i.e. the BR-frame

* Poznań University of Technology, Institute of Electronics and Telecommunication, Piotrowo 3A 60-965 Poznań, Poland, e-mail: [lblaszak, domanski, smack]@et.put.poznan.pl, tel.: +48 61 6652762, fax: +48 61 6652572.

was used as a temporal reference for the neighboring BE-frames (Fig. 1).

In the experiments, in the high resolution video sequence the number of B-frames between two consecutive I- or P-frames is even.

In the enhancement layer there also exist PI-frames, i.e. frames which are encoded without motion vectors.

The GOP structure version without B-frames is presented in Fig. 3. For the case of temporal decimation with factor 3, another GOP structure is appropriate (Fig. 2).

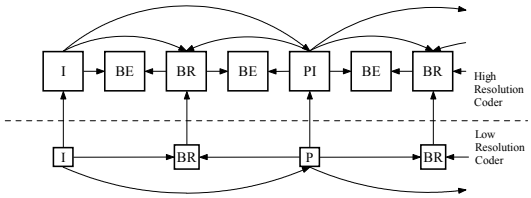

Fig. 1. Selected GOP structure with P- and B-frames in low and high resolution bitstreams.

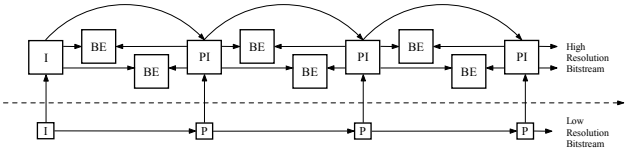

Fig. 2. GOP structure with temporal resolution decimation by factor 3.

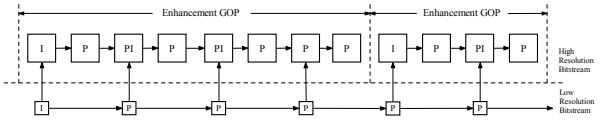

Fig. 3. GOP structure with only P-frames in low and high resolution layers.

3 CODER STRUCTURE

The proposed scalable coder consists of two (or more) motion-compensated coders (Fig. 4) that encode a video sequence and produce two bitstreams corresponding to two different levels of both spatial and temporal resolution. Each of the coders has its own prediction loop with own motion estimation. Therefore the coder produces a bitstream that consists of four major parts:

- encoded transform coefficients for the low-resolution base layer,
- encoded transform coefficients for the high-resolution enhancement layer,
- motion vectors for the low-resolution base layer (mv_l),
- motion vectors for the high-resolution enhancement layer (mv_h).

The proposed scalable encoder was tested in three video coding systems: H.263, MPEG-2 and AVC. Therefore the low-resolution sub-coder was implemented appropriately as a motion-compensated hybrid MPEG-2 encoder of the Main Profile@Main Level (MPEG-2 MP@ML), a motion-compensated hybrid H.263 encoder and a motion-compensated hybrid AVC encoder. The low-resolution sub-coder is a fully compatible with a standard encoder.

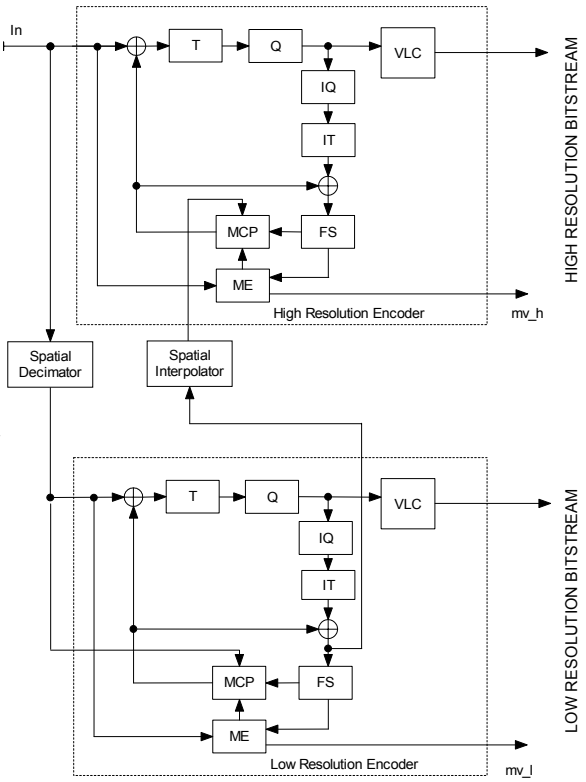

Fig.4 General coder structure

The high-resolution sub-coder is a modification of the appropriate encoder. In the enhancement-layer high-resolution sub-coder additional reference frames can be used for both backward and forward prediction, i.e. interpolated frame from the current low-resolution base-layer frame and linear combinations (averages) of the current interpolated frame and temporal reference. For the latter, independent motion estimation can be performed aiming at estimation of the optimum motion vectors

that yield the minimum prediction error for the reference being an average of spatial and temporal references.

As an extension to the standard compression technique (MPEG-2 and H.263), in the prediction those B-frames which correspond to B-frames from the base layer can be used as reference frames for predicting other B-frames in the enhancement-layer.

In the enhancement-layer layer, some minor modifications of the bitstream semantics are proposed.

In the MPEG-2 standard, the mode of prediction is indicated by the *macroblock_type* which is variable length encoded. The prediction in the enhancement-layer requires transmitting an additional bit per macroblock to identify the selected mode of prediction. Since the enhancement layer is not fully compatible with the MPEG-2 standard, an additional bit was inserted in the macroblock header in the syntax of the enhancement layer bitstream.

In order to decrease the number of bits of the *macroblock_type* indicator, a new variable length codes have been calculated.

Application of the additional reference frames in prediction does not require bitstream syntax modifications and just minor modifications of the semantics for the reference frame variables in sub-coder of the enhancement-layer based on AVC.

The characteristic feature of this structure is independent motion estimation in both sub-coders resulting in optimum motion vectors estimated for both resolution levels. These motion vectors allow exact motion-compensated prediction in both layers.

4 EXPERIMENTAL RESULTS

Three basic series of experiments have been performed for constant quality coding, corresponding to approximately 800kbps for non-scalable H.263, 5Mbps for non–scalable MPEG–2 coding of SDTV signals and various bitrates for non–scalable AVC coding of CIF signal.

In order to evaluate compression efficiency, a verification model of the scalable encoder and an implementation of the MPEG-2 encoder have been used. The base layer encoder is the standard MPEG-2 encoder that processes video in the SIF format. The enhancement layer is characterized by full television resolution (BT.601). The results obtained for the range of few megabits per second are quite promising as the bitrate overhead due to scalability is mostly below 10% with an MPEG-2 reference encoder. In these experiments, we used the sequences of structure shown in Fig. 1.

In order to test the coding performance of the scalable H.263 codec, a series of experiments have been performed with (352 x 288)-pixel sequences. For the H.263 reference codec, the bitrate overhead due to scalability has been measured relative to nonscalable bitstream. This relative overhead depends strongly on the options switched on and on the quality of motion estimation as well. For example, the results for full-pel motion estimation exhibit sometimes even negative overhead for constant bitrate mode in both layers independently controlled.

For the AVC reference codec, the scalable test model has been implemented on the top of standard JVT software version 2.1. In order to test the coding performance of the scalable AVC codec, a series of experiments have been performed with (352 x 288)-pixel sequences.

In the experiments, the following modes have been switched on:

- CABAC coder,
- ¼-pel motion estimation in both layers,
- all prediction modes.

For all the cases above considered, the base-layer was about 30- 40% of the total bitrate for all three decimation factors (temporal, horizontal and vertical) set to 2.

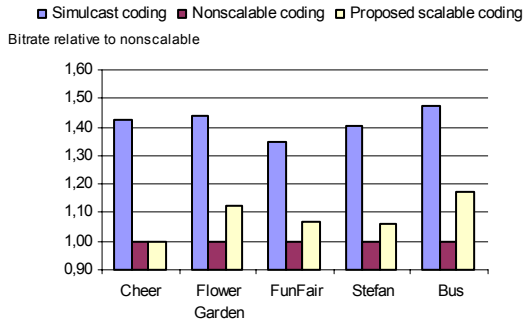

Fig. 5. Approximate bitrate comparison for scalable, nonscalable (single-layer) and simulcast coding at 5 Mbps for non–scalable MPEG–2 coding of SDTV signal.

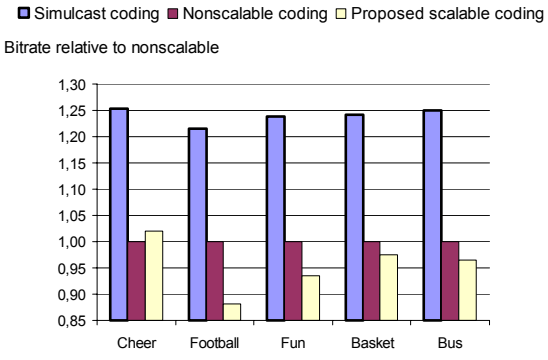

Fig. 6. Approximate bitrate comparison for scalable, nonscalable (single-layer) and simulcast coding at 800 kbps for non–scalable H.263 coding of CIF signal.

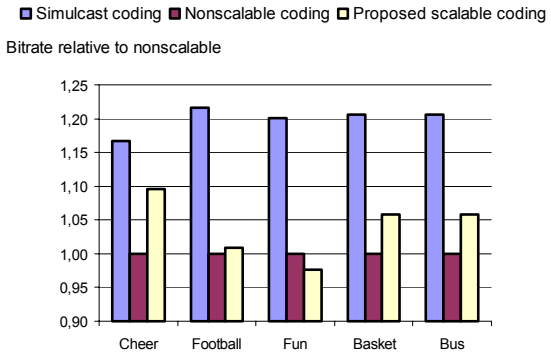

Fig. 7. Approximate bitrate comparison for scalable, nonscalable (single-layer) and simulcast coding at for non–scalable AVC coding of CIF signal.

5 CONCLUSIONS

Described is a scalable extension of the video codec. The basic features of the two-loop coder structure 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.

Acknowledgments

This work has been supported by the Polish National Committee for Scientific Research under a research project in years 2003-2005.

References

[1] ISO/IEC IS 13818-2 / ITU-T Rec. H.262, “Generic coding of moving pictures and associated audio, part 2: video”, November 1994.

[2] Haskell B.G., Puri A., Netravali A.N., Digital video: an introduction to MPEG-2, New York, Chapman & Hall, September 1996.

[3] ISO/IEC/SC29/WG11/MPEG02/N4920, ISO/IEC 14496-10 AVC | ITU-T Rec. H.264, “Text of final committee draft of joint video specification”, Klagenfurt, July 2002.

[4] 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.

[5] ITU-T, “Video coding for low bitrate communication”, Recommendation H.263, 1998.

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

[7] Maćkowiak S., “Scalable Coding of Digital Video”, Doctoral dissertation, Poznań University of Technology, Poznań 2002.

[8] Taubman D. and Zakhor A., “Multirate 3-D subband coding of video,” IEEE Trans. Circ. Syst. Video Techn., vol. 3, pp. 572-588, September 1994.

[9] Kim B.-J., Xiong Z., Pearlman W.A., “Low Bit-Rate Scalable Video Coding with 3-D Set Partitioning in Hierarchical Trees (3-D SPIHT)”, IEEE Transactions on Circuits and Systems for Video Technology, Volume: 10, No. 8, December 2000.

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

[11] Domański M., Łuczak A., Maćkowiak S., „Spatio-Temporal Scalability for MPEG” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 7, pp. 1088-1093, October 2000.

[12] Domański M., Łuczak A., Maćkowiak S., "On improving MPEG Spatial Scalability", Proceedings of the IEEE International Conference on Image Processing ICIP’2000, Vancouver, pp. II-848 - II-851, 2000.

[13] Domański M., Maćkowiak S., "Modified MPEG-2 Video Coders with Efficient Multi-Layer Scalability", Proceedings of the IEEE International Conference on Image Processing ICIP’2001, vol. II, pp. 1033-1036, Thessaloniki, 2001.

[13] ISO/IEC IS 14496-2, “Generic coding of audio-visual objects, Part 2: Visual”, December 1998.

[14] ISO/IEC JTC1/SC29/WG11/MPEG02/ N5421, Study of Final Committee Draft of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC) December 2002, Awaji Island.