SCALABILITY OF MODIFIED AVC/H.264 VIDEO CODECS

Łukasz Błaszak, Marek Domański, Rafał Lange, Sławomir Maćkowiak

Institute of Electronics and Telecommunications, Poznań University of Technology, Poznań, Poland
{lblaszak, rlange, domanski, smack} @ et.put.poznan.pl

ABSTRACT

The paper describes a scalable extension of the AVC/H.264 coder. The proposed coder combines spatial and temporal scalability with FGS (Fine Granularity Scalability). The proposed solution introduces minor modifications of the bitstream semantics and syntax. The coder consists of two independently motion-compensated sub-coders that encode a video sequence and produce two bitstreams corresponding to two different levels of spatial and temporal resolution. The system employs adaptive interpolation as well as luminance-assisted interpolation of chrominance. The functionality of FGS is related to some drift in the enhancement layer. This drift can be limited by excluding temporal prediction in some enhancement layer frames.

1. INTRODUCTION

The state of the art in video compression has just experienced a revolution with the new standard H.264/MPEG-4 AVC/MPEG-4 Part 10. Version 1 of the new video coding standard AVC/H.264 has been already developed [1,2]. H.264/AVC offers significant improvement of coding efficiency as compared to older compression standards such as MPEG-2 [3].

Many applications could be offered with more efficient scalability. The functionality of scalability of video coding is related to the ability to achieve a video of more than one resolution or quality simultaneously from one stream. The existing video compression standards MPEG-2 [3] and MPEG-4 [4] define scalable profiles that exploit classic Discrete Cosine Transform-(DCT)-based schemes with motion compensation. Unfortunately, as proposed by the MPEG-2 and MPEG-4 coding standards, spatial scalability is inefficient because of substantial bitrate overhead as compared to nonscalable schemes. Additionally, the solutions defined in MPEG-2 do not allow flexible allocation of the bitrate.

There exists urgent need to develop efficient algorithms for scalable video coding compliant with the new technology standardized within AVC/H.264. There were many proposals to improve the efficiency of scalable coding [e.g. 5-9]. The goal of this paper is to propose such extensions of AVC/H.264 technology. This proposal extends some ideas recently proposed for MPEG-2 and H.263 [12-13]. The paper reports new original results for modified AVC/H.264 video codecs with spatial, temporal and SNR scalability. The solutions extend the preliminary results which have been already mentioned in [10,11,14].

In the paper, the assumption is to introduce possibly minor modifications of the bitstream semantics and syntax as well as to avoid as much as possible the technologies that are not present in the existing structure of the AVC codec. In particular, it is assumed that the low-resolution base layer bitstream is fully compatible with the AVC/H.264 standard. Moreover, the bitstream syntax is standard, and minor semantics modifications are proposed for the enhancement layer only.

2. GENERAL CODER STRUCTURE

The scalable coder proposed consists of some motion-compensated sub-coders that encode a video sequence and produce bitstreams corresponding to different levels of spatio-temporal resolutions. For example, a three-layer video representation may be produced by a three-loop video coder.

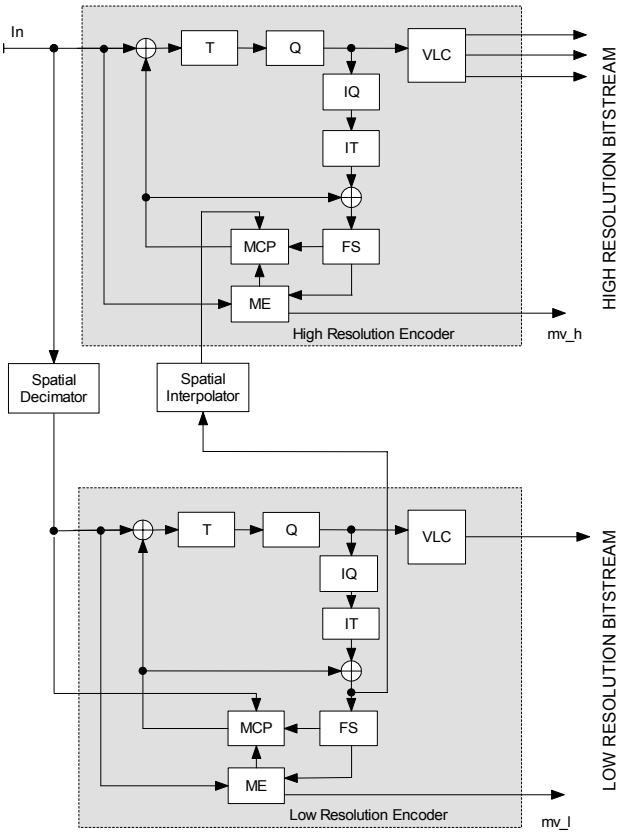
For in-depth analysis, a two-loop encoder has been chosen for considerations. In this case, a coder consists of two motion-compensated sub-coders (Fig. 1). Each of the sub-coders has its own prediction loop with independent motion estimation and compensation. Data partitioning is used in order to obtain the Fine Granularity Scalability (FGS).

The low-resolution sub-coder is implemented as a standard motion-compensated hybrid AVC coder that produces a bitstream with fully standard AVC syntax. The high-resolution sub-coder is a modified AVC coder that is able to exploit the interpolated macroblocks from the decoded base-layer bitstream. These interpolated macroblocks are used as reference macroblocks for prediction whenever they provide lower cost than temporal references. Other additional reference macroblocks are created by averaging the reference of temporal prediction and the interpolated macroblock.

The choice of the spatial decimation filter trades off between high aliasing attenuation and short temporal

response. The results of experimental comparisons prove the importance of the careful choice of the decimation-interpolation scheme.

The system considered employs edge-adaptive bi-cubic interpolation as described in [15]. The technique is applicable to both luminance and chrominance.



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

Fig.1. The structure of the encoder (temporal subsampling is not included in this figure). *VLC – variable-length coder*. *mv_l* and *mv_h* denote motion vectors from the low-resolution and the full-resolution layer, respectively.

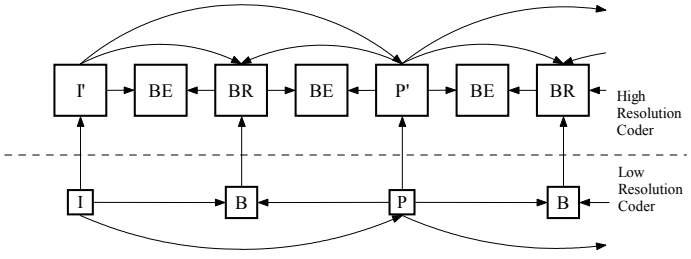


Fig. 2. Exemplary structure of the video sequence.

Here, temporal down-sampling is achieved by partitioning the stream of B-frames: e.g. every second frame is skipped in the low-resolution layer (Fig. 2.). Therefore two kinds of B-frames are obtained: BE-frames that exist in the high resolution sequence only, and BR-frames that exist in both sequences.

4. PREDICTION MODES

In the enhancement layer, the coding scheme takes advantage of additional reference frames being frames interpolated from the decoded current base-layer low-resolution frame. AVC already allows multiple reference frames, and introduction of the above mentioned additional reference frames does not require bitstream syntax modifications and just minor modifications of the semantics for the reference frame variables. The reach AVC bitsream syntax may be used to represent data resulting from various prediction modes with various reference frames, in particular, data resulting from:

- spatial interpolation from the low-spatial-resolution layer,
- an average of the latter and a temporal prediction.

Moreover, for the averaged prediction mode, 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. This option was used in the experiments reported further. The new prediction modes are carefully embedded into the mode hierarchy of the AVC coder thus obtaining the binary codes that correspond to the mode probabilities.

The base layer bitstream is fully compliant with the standard single-layer AVC bitstream syntax and semantics. In the enhancement layer with higher spatial resolution, some minor modifications of the bitstream semantics are needed.

5. EXPERIMENTAL RESULTS

The scalable test model for coder and decoder has been implemented on the top of standard AVC software. As compared to [10], the above described improved prediction was used. In order to test the coding performance of the scalable AVC codec with the averaging mode 2, a series of experiments have been performed with CIF sequences. Horizontal, vertical and temporal subsampling factors have been set to 2 and the video sequence structure was that from Fig. 2.

In the experiments, the following modes have been switched on: CABAC coder, ¼-pel motion estimation in both layers, all prediction modes. The values of the quantization parameter were defined independently for I-frames (QP_I), P-frames (QP_P) and B-frames (QP_B). Within

a scalable coder, the base layer bitrate was about 15% to 22% of the total bitrate produced by a scalable coder for both layers.

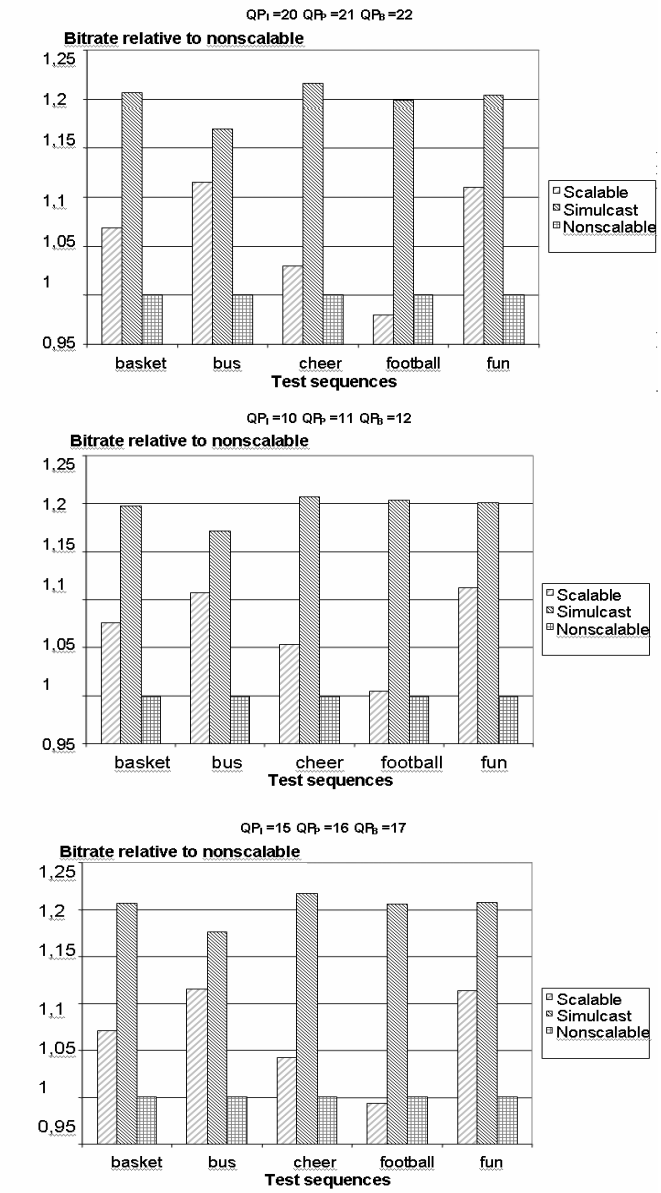


Fig. 3. Approximate bitrate comparison for scalable (two-layer), nonscalable (single-layer) and simulcast coding.

For such test conditions, the approximate bitrate overhead due to scalability was between -1% and 12% of the bitrate for the nonscalable (single-layer) codec. For almost all cases, scalable coder performed better than simulcast coding. Usually scalable coding performance was substantially higher than that of simulcast (Figs. 3,4).

Fine granularity scalability (FGS) is obtained via macroblock partitioning. Except from headers and motion vectors, the bitstreams can be arbitrarily split into layers and multi-layer fine granularity can be achieved (Fig.5). The drawback of this strategy is accumulation of drift. Fortunately, drift propagation is limited by insertion of I-frames into the enhancement layer (often related to GOP structures). Such enhancement-layer I-frames are encoded using less numbers of bits than single-layer I-frames. It is because the bitstream syntax of these frames is that of P-frames but with no motion vectors and with the interpolated base-layer frames used as reference frames. Furthermore, drift in the full resolution part may be reduced by more extensive use of the low resolution images as reference.

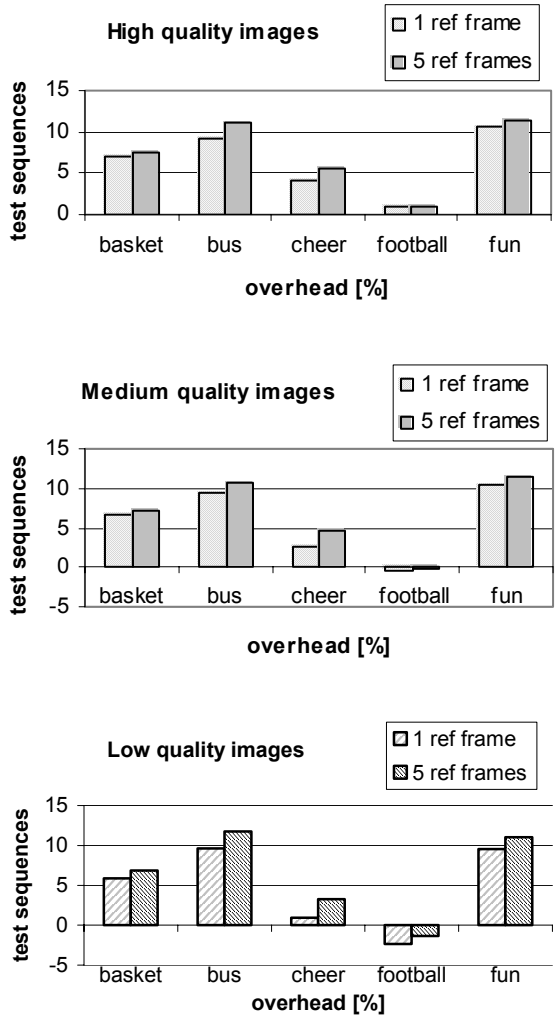


Fig. 4. Bitrate overhead (with respect to single layer coding) for scalable coding exploiting 1 and 5 temporal reference frames.

The motion vectors are encoded independently in the both layers. The authors have tested median predictions from the base-layer vectors (Fig. 6). Unfortunately, no improvement was obtained in most cases.

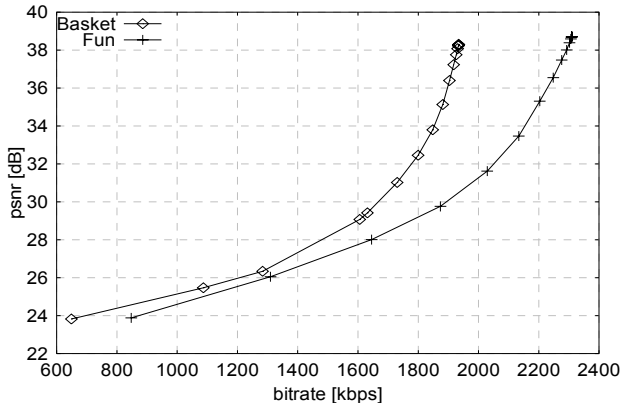


Fig. 5. Rate-distortion curves for FGS in the extended AVC codec: test sequences *Fun* and *Basket.*

7. CONCLUSIONS

Described is a scalable extension of the AVC codec. The basic features of the multi-loop coder structure are: mixed spatio-temporal scalability, and independent motion estimation for each motion-compensation loop, i.e. for each spatio-temporal resolution layer.

The scalable coder exhibits good satisfactory coding performance. For the two-layer system with spatio-temporal scalability, the bitrate overhead due to scalability varies between 0% and 30% depending on sequence content and bitrate allocation. For almost all cases, scalable coder performed better than simulcast coding. Usually scalable coding performance was substantially higher than that of simulcast.

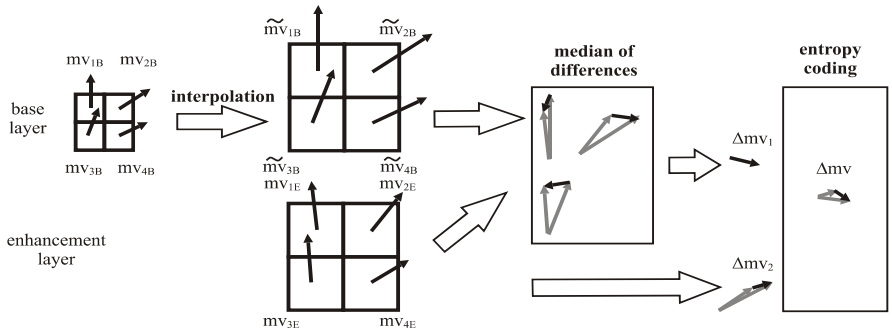


Fig. 6. Motion vector prediction using low-resolution vectors from the base layer.

Scalable coder complexity is similar to that of the simulcast structure. The major additional operations are spatial interpolation and additional mode selection. More complex is an optional additional motion estimation for averages of spatial and temporal reference operation that improves coding efficiency.

ACKNOWLEDGEMENT

The work has been supported by Polish Committee for Scientific Research.

REFERENCES

[1] ISO/IEC/SC29/WG11/MPEG02/N5555, FDIS 14496-10 March 2003..

[2] IEEE Trans. Circuits and Syst. . Video Techn. Special Issue on H.264/AVC Video Coding Standard, vol. 13, July 2003.

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

[4] ISO/IEC IS 14496-2, “Generic coding of audio-visual objects, part 2: visual” 1998.

[5] S. Maćkowiak, “Scalable Coding of Digital Video”, Doctoral diss., Poznań Univ. of Technology, Poznań 2002.

[6] S. Maćkowiak, “Multi-Loop Scalable MPEG-2 Video Coders”, Springer Verlag Lecture Notes in Computer Science, vol. 2756, pp. 262-269, 2003.

[7] U. Benzler, "Scalable multi-resolution video coding using a combined subband-DCT approach", Picture Coding Symp., Portland, USA, pp. 21-23, April 1999.

[8] Y.-F. Hsu, C.-H. Hsieh, Y.-C. Chen, “Embedded SNR scalable MPEG-2 video encoder and its associated error resilience decoding procedures” Signal Processing: Image Communication 14 (1999) pp. 397-412, 1999.

[9] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "On Video SNR Scalability", Proc. 1998 IEEE International Conf. on Image Processing, vol. 3, pp. 934-936, Chicago, 1998.

[10] Ł. Błaszak, M. Domański, S. Maćkowiak, "Spatio-Temporal Scalability in AVC codecs", Doc. MPEG2003/M9469 Pattaya, March 2003.

[11] Ł. Błaszak, M. Domański, “Modified AVC Codecs with Spatial and Temporal Scalability”, Doc. MPEG2003/ M9895, Trondheim, July 2003.

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

[13] M. Domański, A. Łuczak, S. Maćkowiak, “Spatio-temporal scalability for MPEG video coding”, IEEE Trans. Circ. and Syst. Video Technol., vol. 10, pp. 1088-1093, Oct. 2000.

[14] M. Domański, Ł. Błaszak, S. Maćkowiak, "AVC Video Coders with Spatial and Temporal Scalability", Picture Coding Symp., pp. 41-46, Saint Malo, 2003.

[15] G. Ramponi, Warped distance for space-variant linear image interpolation, IEEE Transactions on Image Processing, vol. 8, pp. 629-639, May 1999.