SIMPLE AVC-BASED CODECS WITH SPATIAL SCALABILITY

Rafał Lange, Łukasz Błaszak, Marek Domański

Poznań University of Technology, Institute of Electronics and Telecommunications, Poznań, Poland

ABSTRACT

Proposed is a coder that produces a layered video representation with layers corresponding to different spatial resolutions. A coder consists of AVC-based sub-coders with independent motion estimation and compensation. Improved motion-vector encoding is provided for the enhancement layer. Other codec features include adaptive interpolation from the base layer, full AVC-compatibility of the base layer, and syntax compatibility of the enhancement-layer bitstream. Interpolated reconstructed base-layer frames are used as additional reference frames in the enhancement layer. The respective prediction modes are embedded into prediction strategy of AVC. Experimental results prove that improved motion-vector encoding results in bitrate reduction of up to 13% of the enhancement-layer bitrate for motion vectors. As measured for all layers together, compression efficiency is significantly higher than for simulcast, but scalable codec complexity is only slightly higher than complexity of the respective simulcast codec.

1. INTRODUCTION

Recently, video compression has experienced substantial progress with introduction of the new version of hybrid video coding called Advanced Video Coding (AVC). This technology is just being standardized by ISO as part 10 of MPEG-4, and by ITU as Recommendation H.264 [1,2]. Currently, AVC is considered in context of many applications, including replacement of MPEG-2 [3]. Unfortunately, AVC Version 1 video codec does not support scalability that is currently considered as an important functionality for many applications, e.g. wireless systems with bandwidth variations and fadings, video broadcasting in heterogeneous communication networks, unequal error protection etc. [4]. Importance of scalability has been recognized also by MPEG that issued requirements and call for proposals for a technology for scalable coding of video [5]. There exist already several technologies that can be used. After success of scalable compression of still images, wavelet-based techniques are considered as good candidates. Among them, very promising are techniques that exploit three-dimensional spatio-temporal analysis with motion-compensated filtering [6-9]. They are not compatible with AVC and exhibit lower efficiency in low-latency modes. Nevertheless new proposals [10] seem to make this limitation less critical. On the other hand, interesting AVC-compliant techniques have been proposed recently, mostly for SNR-scalability with fine granularity [11-15]. For spatial and temporal scalability, there were proposals similar to those earlier proposed in order to improve scalability for MPEG-2 [16,17]. Here, we are going to propose a simple technique for spatial scalability that extends earlier proposals for AVC-based spatial and temporal scalability [18,19].

Proposed is a coder that produces a layered video representation with layers corresponding to different spatial resolutions. The very basic idea is taken from old solutions already proposed for MPEG-2 [3], i.e. a spatially scalable coder consists of sub-coders corresponding to individual layers representing various resolutions of pictures. This solution was not practical as proposed in MPEG-2 standard [3]. The reason was low compression efficiency. Nevertheless its improvements resulted in higher coding efficiency [16,17]. Similar approach has been already applied successfully to hybrid spatial and temporal scalability on AVC platform [18,19].

Here, only spatial scalability is considered but with some improvements in respect to earlier solutions from [18,19]. The most important improvement is related to better encoding of motion vectors in the enhancement layer.

2. CODER STRUCTURE

As proposed, a scalable coder consists of sub-coders. The base-layer sub-coder produces a bitstream that represents video with the lowest spatial resolution. Higher-resolution video is obtainable from the higher-layer bitstream together with the base-layer one. For the sake of simplicity, a two-layer system (Fig. 1) will be considered. Its generalization to more layers is straightforward and the authors have already tested it experimentally with success.

The following are the characteristic features of the coder proposed:

- The base layer is fully AVC-compliant.
- Standard bitstream syntax is preserved in all layers with minor semantics modifications in the enhancement layer.
- Interpolated reconstructed base-layer frame is used as an additional reference frame in the enhancement layer.
- Edge-adaptive interpolation from the base layer is used.
- There exists an additional prediction mode using the interpolated base-layer frame averaged with the result of temporal prediction.
- Independent motion estimation and compensation is used in individual sub-coders for individual layers.
- Improved motion vector encoding is provided for the enhancement layer.
- Codec complexity is comparable to the complexity of a pair of codecs used for simulcast coding of two layers.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

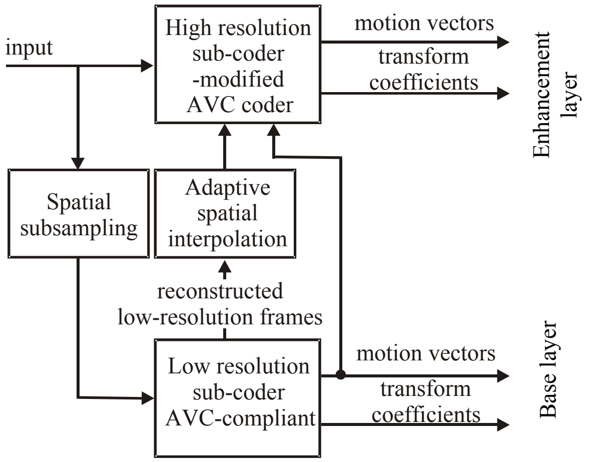

Figure 1: Generic two-layer structure of a spatially scalable coder.

3. IMPROVED ENCODING OF MOTION-VECTORS IN THE ENHANCEMENT LAYER

In a spatially scalable coder, substantial increase of coding efficiency has been obtained by use of improved encoding of motion vectors in the enhancement layer. As mentioned above, motion vectors are estimated independently for each layer. Optimum motion vectors are used in both layers, thus resulting in near-minimum data bitrates of transform coefficients for prediction residuals. In previous proposals [18,19], both motion-vector bitstreams were being encoded independently using standard AVC technique that exploits motion vector prediction using the neighboring motion vectors from the same layer. Here, in the enhancement layer, proposed is prediction that uses also a co-located motion vector from the base layer. As intra-layer prediction of motion vectors is more exact than prediction from the base layer, motion vector prediction from the base layer is used only in the cases when there is no adequate reference motion vector, i.e. a neighboring block is intra-coded or uses another temporal reference frame. Because of lower spatial resolution of pictures in the base layer, the base layer motion vector has to be scaled when used for prediction in the enhancement layer.

In the enhancement layer, the modified-AVC median prediction exploits motion vectors from neighboring blocks A, B, C and D, and from the base layer (X) (Fig. 2). The respective prediction scheme (Table 1) has been obtained by modification of the standard AVC scheme for the cases where no good reference motion vector is available. In the base layer, motion vector prediction is fully standard. In both layers, the prediction residuals are encoded in a standard way, i.e. using efficient context-adaptive binary coding (CAVLC or CABAC) [1,2].

The AVC standard defines a directional prediction for motion vectors of blocks of size 8×16. For such cases, directional prediction is used both in the base and in the enhancement layer in the standard way (Fig. 3).

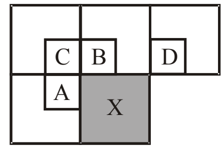

Figure 2: Median motion vector prediction.

Table 1: Motion vector prediction scheme.

A	B	C	D	X	prediction
a	a	a			median(A,B,C)
a	a	0	a		median(A,B,D)
a	a	b,0	b,0	a,b,0	median(A,B,X)
a	b,0	a		a,b,0	median(A,C,X)
a	b,0	0	a	a,b,0	median(A,D,X)
b,0	a	a		a,b,0	median(B,C,X)
b,0	a	b,0	a	a,b,0	median(B,C,X)
a	b,0	b,0	b,0	b,0	A
a	b,0	b,0	b,0	a	X
b,0	a	b,0	b,0	b,0	B
b,0	a	b,0	b,0	a	X
b,0	b,0	a		b,0	C
b,0	b,0	a		a	X
b,0	b,0	0	a	b,0	D
b,0	b,0	0	a	a	X
b,0	b,0	b,0	b,0	a	X
b,0	b,0	b,0	b,0	b,0	X

a – available (the same reference frame),
b – available (different reference frame or intra coded),
0 – unavailable.

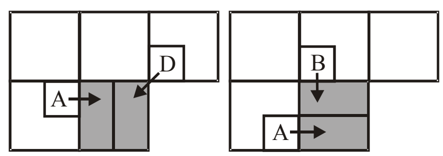

Figure 3: Directional motion vector prediction.

4. PREDICTION IN THE ENHANCEMENT LAYER

In the enhancement layer, the coding scheme takes advantage of two additional reference frames: the frame interpolated from the decoded current base-layer low-resolution frame, and an average of the latter and the last temporal reference frame. Available could be even more reference frames obtained as combinations of the interpolated frame and various temporal references. Nevertheless those possibilities have not been exploited in the experiments reported in this paper.

The enhancement-layer sub-coder employs additional prediction mode that exploits the current interpolated base-layer frame as the reference. Another mode exploit averages of temporal prediction and spatial interpolation as references. The results from [19] are applicable for binary encoding of additional prediction modes.

Scalable coding is efficient if temporal prediction and prediction from the base layer are mixed with substantial probability for each mode. In the enhancement layer, poor interpolation would result in choice of temporal prediction only.

It would mean that scalable coder works like simulcast arrangement. Therefore efficient but simply implementable edge-adaptive interpolation technique from [20] has been used in the test model.

This technique of edge-adaptive interpolation is an extension of the standard non-adaptive bi-cubic separable interpolation that can be described as follows. The two-dimensional interpolation is performed in two steps: horizontal and vertical. Let $f(x)$ is the value to be estimated, and the nearest available values are located at coordinates x_k (left) and x_{k+1} (right). Let $s = x - x_k$, $1 - s = x_{k+1} - x$, where $0 \leq s \leq 1$.
By bi-cubic separable interpolation, there is

$$f(x) = f(x_{k-1})(-s^3 + 2s^2 - s)/2 + f(x_k)(3s^3 - 5s^2 + 2)/2 + f(x_{k+1})(-3s^3 + 4s^2 + s)/2 + f(x_{k+2})(s^3 - s^2)/2,$$

where x_{k-1}, x_k, x_{k+1} and x_{k+2} are the positions of four neighboring known pixels.

In the edge-adaptive scheme, a modified value s' is used instead of s.

$$s' = s - kAs(s - 1),$$

where k is a positive parameter that controls the intensity of warping and A is a function of asymmetry of the data in the neighborhood of x:

$$A = (|f(x_{k+1}) - f(x_{k-1})| - |f(x_{k+2}) - f(x_k)|)/(L - 1),$$

where $l = 256$ for 8-bit sample representation. In the experiments, it was $k = 3.05$.

Experimental results prove that application of edge-adaptive interpolation improves coding efficiency of the spatially scalable coder significantly, often more than by 1 dB in PSNR of luminance. Edge-adaptive interpolation produces “better” edges in the interpolated reference frames thus increasing probability that interpolated data will be chosen as the best prediction.

5. EXPERIMENTAL RESULTS

In order to examine coding efficiency, scalable codec test model has been built on top of the AVC reference software version 7.3 with CABAC. Here, reported are the results for 4:2:0 CIF (352×288 luminance pixels) test video sequences *Stefan*, *Football* and *Bus*. It means that the base layer corresponds to the QCIF resolution (176×144 pixels) with the same frame rate.

First of all, the new motion vector prediction scheme has been compared to standard AVC-like prediction in both layers. Application of the new scheme results in up to 13% reduction of the enhancement-layer motion-vector data (Fig. 4). This reduction is higher for lower bitrates as higher percentage of video data corresponds to motion vectors for lower bitrates. Moreover this reduction depends also on video content.

In order to test spatially scalable compression efficiency, comparisons to nonscalable and simulcast coding have been performed, i.e. for all three coding scenarios, the bitrates have been compared for the same decoded video quality measured as PSNR for luminance with the full resolution (Figs. 5-7).

In all cases, scalable coding efficiency is higher than for simulcast, i.e. the scalable bitstream exhibits bitrate being smaller by 1-18% than the sum of two independent bitstreams with CIF and QCIF resolutions (Fig. 8). The improvement depends strongly on video content.

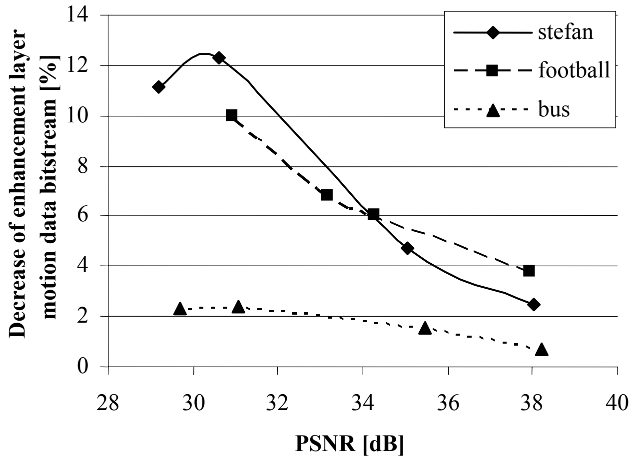

Figure 4: Decrease of the enhancement-layer motion data bitrate due to improved motion vector prediction.

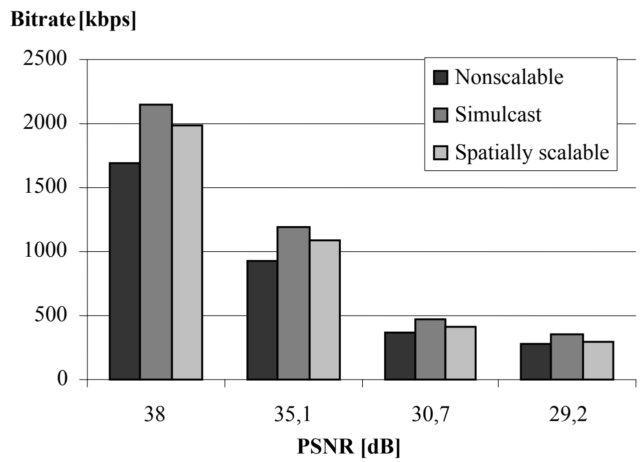

Figure 5: Bitrate comparison for nonscalable, simulcast and scalable coding – test sequence *Stefan*.

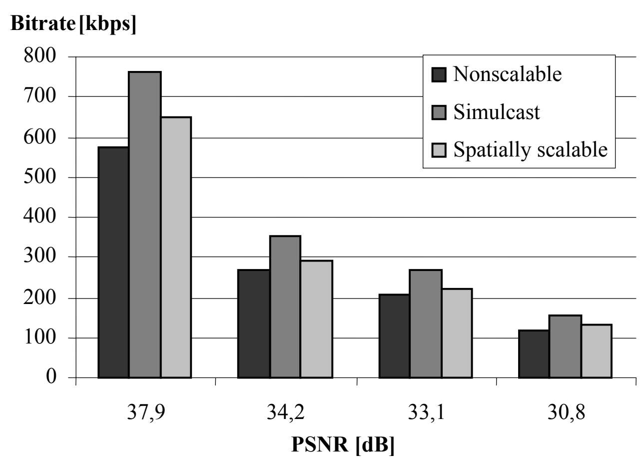

Figure 6: Bitrate comparison for nonscalable, simulcast and scalable coding – test sequence *Football*.

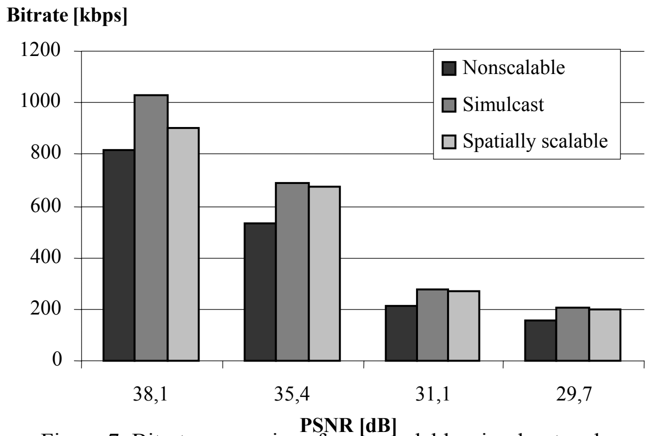

Figure 7: Bitrate comparison for nonscalable, simulcast and scalable coding – test sequence *Bus*.

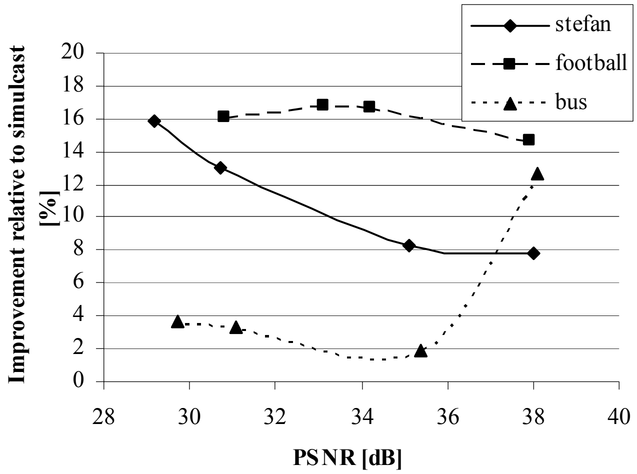

Figure 8: Compression efficiency improvement relative to AVC simulcast – bitrate decrease relative to simulcast.

5. CONCLUSIONS

The paper describes a simple coder with the functionality of spatial scalability. The complexity penalty for scalability is low both in the encoder and in the decoder. It mainly consists of:

- Decimation aimed at production of the low-resolution base layer (at the encoder only).
- Edge-adaptive interpolation that produces an additional reference frame for the enhancement layer (both in the encoder and the decoder).
- More complicated decisions for intra- and inter-frame prediction mode selection (at the encoder only).

The conclusion is that complexity of the scalable codec is not much higher than that of the simulcast arrangement, but coding efficiency is definitely improved.

The codecs are AVC-compatible with only minor enhancement-layer bitstream semantics differences needed to name new prediction modes.

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

REFERENCES

[1] ISO/IEC JTC 1/SC29/WG11/N5555, Text of ISO/IEC FDIS 14496-10: "Information Technology – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding."

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

[3] ISO/IEC IS 13818-2, "Generic Coding of Moving Pictures and Associated Audio, Part 2: Video", Nov. 1994.

[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] ISO/IEC JTC 1/SC29/WG11, Doc. N6025, "Requirements and Applications for Scalable Video Coding," Oct. 2003.

[6] J. R. Ohm, "Three-Dimensional Subband Coding With Motion Compensation," *IEEE Trans. Image Proc.*, vol. 3, pp. 559-571, September 1994.

[7] S.-J. Choi, J. Woods, "Motion Compensated 3-D Subband Coding of Video," *IEEE Trans. Circ. Syst. Video Techn.*, vol. 8, pp.155-167, February 1999.

[8] S. Hsiang, J. Woods, "Embedded Video Coding Using Invertible Motion Compensated 3-D Subband/Wavelet Filter Bank,“ *Signal Proc.: Image Commun.,* vol.16, p.705-724, 2001.

[9] D. Turaga, M. van der Schaar, **"**Unconstrained Motion Compensated Temporal Filtering," Doc. ISO/IEC JTC 1/SC29/WG11 M8388, May 2002.

[10] W.-J. Han, Successive Temporal Approximation and Referencing (STAR) for Improving MCTF in Low End-to-end Delay Scalable Video Coding, Doc. ISO/IEC JTC1/SC29/WG11 M10308, Dec. 2003.

[11] H. Schwarz, D. Marpe, Th. Wiegand, "SNR-scalable Extension of H.264/AVC," Doc. JVT-J035, Dec. 2003.

[12] K. Ugur, P. Nasiopoulos, "Design Issues and Proposal for H.264 Based FGS," Doc. ISO/IEC JTC1/SC29/WG11 M9505, March 2003.

[13] W.-H. Peng, T. Chiang, H.-M. Hang, "Context-Adaptive Binary Arithmetic Coding for Fine Granularity Scalability," Doc. ISO/IEC JTC1/SC29/WG11 M8606, July 2002.

[14] H.-Ch. Huang, Ch.-N. Wang, T. Chiang, "A Robust Fine Granularity Scalability Using Trellis-Based Predictive Leak," *IEEE Trans. Circ. Syst. Video Techn.*, vol. 12, 372-385, 2002.

[15] X. Sun, F. Wu, S. Li, "The Description and Proposed Syntax of JVT-based FGS", Doc. ISO/IEC JTC1/SC29/WG11 M8204, March 2002.

[16] M. Domański, A. Łuczak, S. Maćkowiak, "Spatio-Temporal Scalability for MPEG Video Coding", *IEEE Trans. Circ. Syst. Video Techn.*, vol. 10, pp. 1088-1093, Oct. 2000.

[17] S. Maćkowiak, *Scalable Coding of Digital Video*, Doctoral dissertation, Poznań Univ. Technology, Poznań 2002.

[18] Ł. Błaszak, M. Domański, S. Maćkowiak, Spatio-Temporal Scalability in AVC codecs, Doc. ISO/IEC JTC1/SC29/WG11 M9469, March 2003.

[19] M. Domański, Ł. Błaszak, S. Maćkowiak, "AVC Video Coders with Spatial and Temoral Scalability," *Picture Coding Symposium*, Saint Malo, pp. 41-47, April 2003.

[20] G. Ramponi, Warped Distance for Space-Variant Linear Image Interpolation, *IEEE Trans. Image Proc.*, vol. 8, pp. 629-639, May 1999.