REPRESENTATION OF MOTION VECTORS IN MODIFIED ADVANCED VIDEO CODING WITH SPATIAL SCALABILITY

Rafał Lange, Łukasz Błaszak, Marek Domański

Poznań University of Technology, Institute of Electronics and Telecommunications
Piotrowo 3A, PL-60-965 Poznań, Poland (Europe)
phone: +48 61 665 27 62, email: {rlange, lblaszak, domanski } @et.put.poznan.pl
web: www.multimedia.edu.pl

ABSTRACT

The paper describes improved representation of motion vectors in enhancement-layer bitstream of a spatially-scalable modified AVC video encoder. Proposed solution exploits rescaled motion vectors from the base layer for better predictive encoding of motion vectors in the enhancement layer. The base-layer vectors are used as reference for prediction in the enhancement layer only when no good reference is available in the enhancement layer. The low-resolution base-layer bitstream is fully compatible with AVC standard, while few modifications have been done in the enhancement layer bitstream semantics.

1. INTRODUCTION

Recently, a new hybrid video coding standard AVC (Advanced Video Coding) has been developed [1,2]. The AVC video encoding is considered to be about twice more efficient than its predecessors like MPEG-2. The most important improvements of AVC include: advanced intra-frame prediction, variable size of blocks in inter-frame prediction, multi-hypthesis motion-compensated prediction and sophisticated context-adaptive binary coding (with adaptive arithmetic coding CABAC as an option).

The whole encoded bitstream of a hybrid video coder consists of three major components: control data, motion vectors and transform coefficients of the prediction residual. In AVC, motion vector bitstream is very efficiently compressed by use of predictive coding. Advanced Video Coding represents state-of-art technology of video compression, and this technology will be used as a platform for scalability considerations in this paper.

Scalability of video coding provides an ability to recover video even from some parts of a bitstream [3,4]. Using only a certain part of scalable video bitstream, a receiver is able to reconstruct video but with reduced resolution and/or quality. This feature has been recognized as very important for video communications and broadcasting over heterogeneous or error-prone networks and networks with varying quality of service, e.g. wireless. The current version of the AVC standard does not support scalability but MPEG is considering technologies of scalable video coding for prospective standardization [5]. State-of-art technologies are mostly related either to wavelet or to hybrid video coding. The very promising wavelet-based codecs are inherently scalable but they are not related to commonly used and standardized technology, and many solutions are unable to work in low-latency mode.

In this paper, we consider scalable hybrid codecs built on the AVC platform, i.e. scalable video codecs that exploit codec building blocks and bitstream syntax of AVC. Such a structure has been already proposed for spatio-temporal and FGS scalability [4,6]. For spatio-temporal scalability, from a part of the bitstream, video can be extracted with reduced both spatial and temporal resolutions. In contrary to that, here, we consider problems related to spatial scalability only.

For the sake of simplicity, assume a scalable coder produces a bitstream that consists of two parts:

- base layer bitstream that represents low-resolution video,
- enhancement-layer bitstream that provides information necessary to enhance video to full spatial resolution.

Obviously, further considerations are extendable onto systems that produce multi-layer video representations.

The paper deals with such scalable coders that consist of sub-coders with independent motion estimation for each layer [4,7]. Application of such a coder structure implies presence of motion vectors both in the base and enhancement layers. Therefore, full bitstream contains two sets of motion vectors: sparse motion field for the low-resolution base layer and more dense motion-field for the full-resolution enhancement layer. Moreover both motion fields differ in scale of motion vectors as they represent shifts in frames of different resolutions. Of course, there is a question how to exploit mutual correlation of these two motion fields, i.e. how to compactly represent enhancement motion vectors with use of the information from the low-resolution base layer. This problem is treated in our paper.

The above mentioned problem is somewhat similar to that known from video transcoding [10] where mostly motion vector mapping from higher resolution to lower resolution is considered [8,9], in contrary to our problem.

2. PREDICTIVE CODING OF MOTION VECTORS IN ADVANCED VIDEO CODING

In AVC [1,2], adaptive entropy coding (variable-length or arithmetic coding) is used for prediction residual obtained by prediction of the current motion vector from the neighbouring vectors. There are two predictors defined in AVC standard: median predictor and directional predictor.

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

The directional predictor is used when macroblock is partitioned into two vertical or horizontal partitions of 8×16 luma samples. In all other cases motion vectors components are predicted as a median of three respective components from the neighbouring blocks (Fig.1).

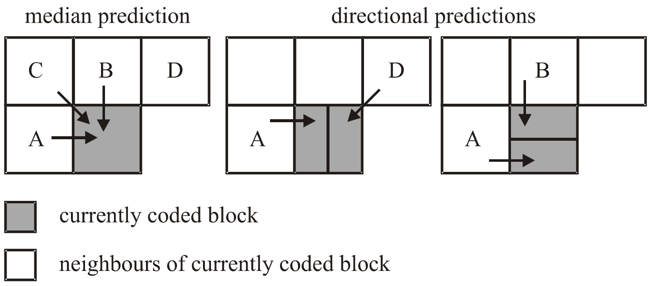

Figure 1: Motion vector prediction schemes used in AVC.

Median prediction is also used for the cases when one of the reference motion vectors is not available because:

- a neighbouring block is outside picture boundaries,
- a neighbouring blocks is intra-frame coded,
- another reference frame is used for motion-compensated prediction in a neighbouring block.

In such cases, unavailable motion vectors are considered to be zero motion vectors. Moreover, when only one reference motion vector is available, this motion vector is directly taken as a prediction. Similar process is invoked for prediction of motion vectors from skipped and “spatially-directly” coded macroblocks while “temporally-directly” coded macroblocks use completely different scheme for motion vector derivation.

It is common experience that the above described prediction scheme proved to be very efficient.

3. SPATIALLY SCALABLE MODIFIED AVC ENCODER

Spatially scalable video encoder consists of two subcoders (Fig. 2) with independent motion prediction and compensation. Such a structure has proved to quite efficient for both spatial and temporal scalability [4,7] and adaptive spatial interpolation applied [4,7].

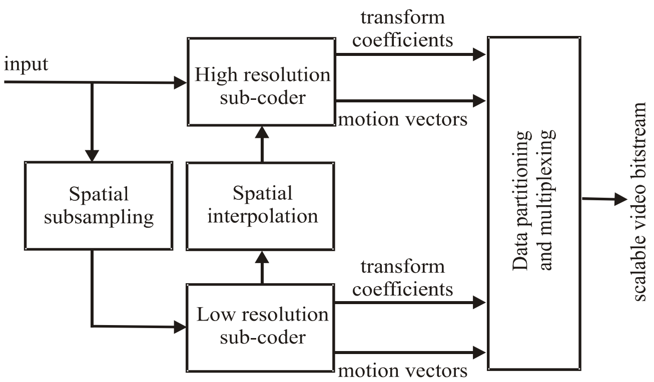

Figure 2: General structure of a scalable coder.

The low resolution sub-coder is fully AVC-compliant while the enhancement-layer (full-resolution) sub-coder is able to exploit adaptively interpolated macroblocks from the decoded base layer, whenever the inter-layer prediction provides lower coding cost [4].

4. PREDICTION OF ENHANCEMENT MOTION VECTORS FROM BASE LAYER

For a spatially scalable coder, enhancement-layer bitrate is usually about 60-75% of the total bitrate. A substantial part of this bitrate corresponds to enhancement-layer motion vectors (Fig.3).

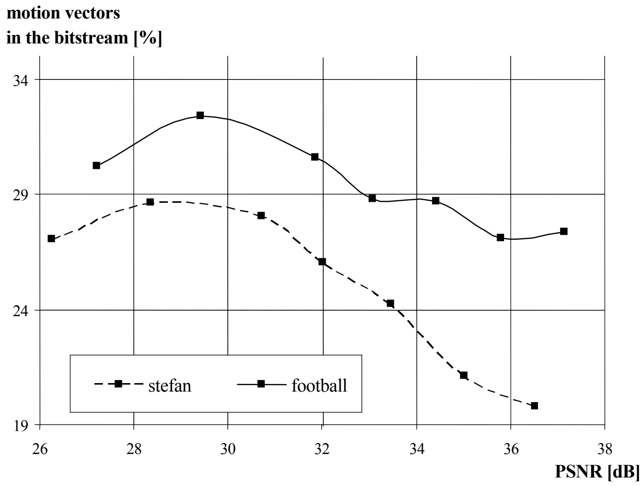

Figure 3: Percentage of motion vectors data in enhancement-layer bitstream of a scalable video coder (test sequences: CIF, 30fps).

A straightforward approach to encode these motion vectors is to exploit predictions from the base-layer motion vectors. The authors made extensive experiments, which proved that in most cases, motion-vector prediction from the base layer is worse than that from the enhancement layer itself. The reasons for that are: higher accuracy of the enhancement motion vectors and local differences between the two motion fields. The latter problem may be solved by smoothing both motion fields. Such smoothed motion fields yield lower motion-vector bitrates due to higher autocorrelation of the motion fields encoded by predictive encoders. Moreover co-located motion vectors are more correlated, i.e. base-layer motion vectors may be more efficiently used for prediction of enhancement-layer motion vectors. The cost for that is increased bitrate for transform coefficients in both layers, due to worse prediction with not optimum motion vectors. Unfortunately, the bitrate increase for transform coefficients was never fully compensated by the bitrate decrease for motion vectors that have been efficiently predicted from the base layer. In many tests, the total bitrate increase was always small and never exceeded 4% [11].

Achieved results proved, that standard AVC-encoding of motion vectors is very efficient and straightforward attempts to improve encoding of motion vectors in spatially scalable video coders may not be successful.

5. IMPROVED CODING OF MOTION VECTORS IN ENHANCEMENT LAYER

The motion vector prediction error is small when neighbouring motion vectors are highly correlated. Unfortunately, experiments show that smoothing motion vector field results in increase of transform-coefficient bitrate [11]. As mentioned in the previous section, neighbouring motion vectors from the enhancement layer are generally better suited as references for prediction than the co-located rescaled motion vectors from the base layer. Nevertheless prediction of a motion vector may be improved in the cases when the reference motion vectors are unavailable because:

- a neighbouring block is outside picture boundaries,
- a neighbouring block is intra-frame coded,
- another reference frame is used for motion-compensated prediction in a neighbouring block.

In all above cases, standard predictor assumes zero value for the reference vector or it does not perform median prediction at all, taking directly a single motion vector as a prediction instead. In these cases, motion vector prediction is poor. The remedy for that is to exploit a rescaled motion vector from the co-located corresponding block from the base-layer. The respective rescaled base-layer motion vector may be used as reference for median or directional prediction (Fig. 4).

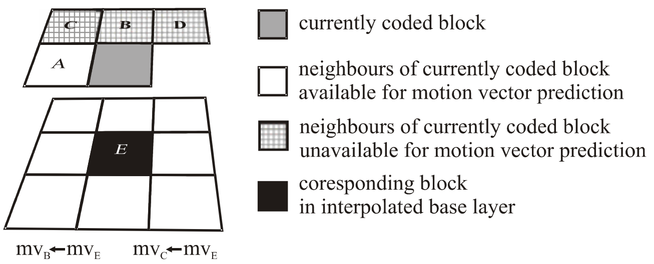

Figure 4: Exploitation of motion vectors from the base layer by motion vector prediction in the enhancement layer.

It is very important to notice that the quality of prediction improves, because an alternative motion vector from the base layer is used for prediction only when no proper motion vector from the enhancement layer is available.

In this way prediction of motion vectors is improved, i.e. prediction residual is reduced, and as a consequence, reduced is the number of bits needed to encode motion vectors.

6. EXPERIMENTAL RESULTS

In order to test efficiency of the improved compressed representation of motion vectors in spatially scalable AVC encoder, a series of experiments have been performed with CIF (352×288) video test sequences. For the same video quality (PSNR for luminance) bitrates have been compared for the standard and proposed coding schemes.

Table 1: Experimental results for *football* sequence.

		motion vectors representation							
		standard	improved	standard	improved	standard	improved	standard	improved
base layer	bitrate [kbit/s]	188		84		63		36	
	PSNR [dB]	37		33		32		30	
enhancement layer	bitrate [kbit/s]	475	462	218	209	167	160	100	94
	PSNR [dB]	38	38	34	34	33	33	31	31
	bitrate decrease [%]	-	**2.7**	-	**3.8**	-	**4.3**	-	**6.2**
total	bitrate [kbit/s]	663	650	302	293	231	223	137	131
	bitrate decrease [%]	-	**2**	-	**2.8**	-	**3.1**	-	**4.5**

Table 2: Experimental results for *bus* sequence.

		motion vectors representation							
		standard	improved	standard	improved	standard	improved	standard	improved
base layer	bitrate [kbit/s]	208		156		63		46	
	PSNR [dB]	36		35		30		29	
enhancement layer	bitrate [kbit/s]	691	690	518	516	209	206	153	151
	PSNR [dB]	38	38	35	35	34	34	30	30
	bitrate decrease [%]	-	**0. 2**	-	**0.3**	-	**1.2**	-	**1.3**
total	bitrate [kbit/s]	899	898	674	673	271	269	199	197
	bitrate decrease [%]	-	**0.1**	-	**0.2**	-	**0.9**	-	**1**

Table 3: Experimental results for *stefan* sequence

		motion vectors representation							
		standard	improved	standard	improved	standard	improved	standard	improved
base layer	bitrate [kbit/s]	460		267		107		77	
	PSNR [dB]	37		34		29		28	
enhancement layer	bitrate [kbit/s]	1544	1517	845	824	328	307	240	224
	PSNR [dB]	38	38	35	35	31	31	29	29
	bitrate decrease [%]	-	**1.8**	-	**2.4**	-	**6.5**	-	**6.8**
total	bitrate [kbit/s]	2005	1977	1112	1091	435	414	318	301
	bitrate decrease [%]	-	**1.4**	-	**1.9**	-	**4.9**	-	**5.2**

For all tests, the following encoder modes have been switched on:

- CABAC entropy coder,
- enabled all inter-frame prediction modes,
- 5 reference frames,
- only first frame coded in intra-prediction mode.

Three test sequences (199 frames of *football.cif, bus.cif* and *stefan.cif*) have been used. Test results are shown in Tables 1, 2 and 3. Decrease of the enhancement-layer bitrate, when using improved motion vector prediction, is depicted on Fig. 5. Fig.6 shows decrease of motion data (motion vectors and reference frame indices) in a bitstream.

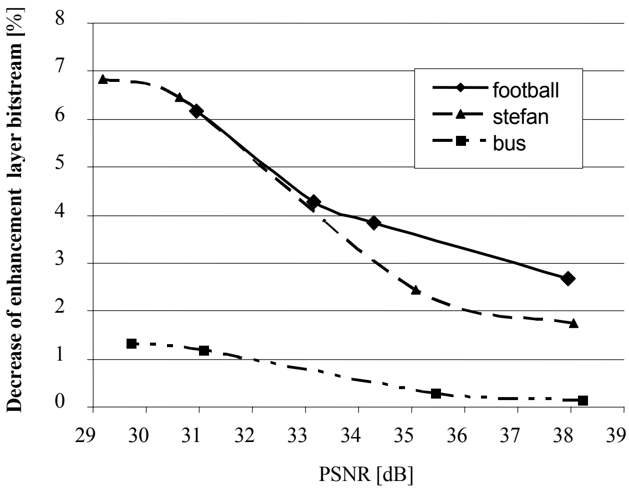

Figure 5: Decrease of the overall enhancement-layer bitrate due to improved motion vector prediction.

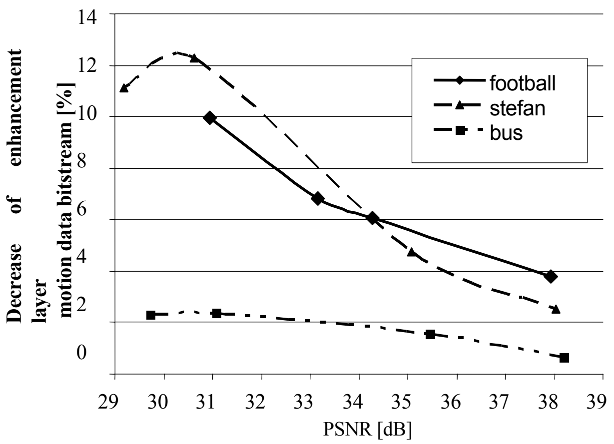

Figure 6: Decrease of the enhancement-layer motion data bitrate due to improved motion vector prediction.

7. CONCLUSIONS

Efficient spatially scalable hybrid video coders have been considered. These coders consist of sub-coders that are independently motion-compensated. Moreover adaptive spatial interpolation is used. These two improvements have been already described [4] and this description was skip out for the sake of brevity of this paper. This paper is focused on another new improvement in scalable hybrid video coders, i.e. on improved enhancement-layer motion vector encoding. This new technique exploits co-located rescaled motion vectors from the base layer only to replace poor reference vectors in the enhancement layer.

The discussion on straightforward joint coding of the base- and enhancement-layer motion vectors is reported briefly in order to show the background of the technique proposed. This proposed technique exploits prediction using co-located rescaled motion vectors from the base layer as median prediction reference in all cases where no good reference exists in the enhancement layer. Experimental results prove that this new technique is always more efficient than independent encoding of motion vectors in two layers of a spatially scalable hybrid video codec based on AVC/H.264 technology.

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

[4] M. Domański, Ł. Błaszak, S. Maćkowiak, "AVC video coders with spatial and temoral scalability," *Picture Coding Symposium*, Saint Malo, April 2003, pp. 41-47.

[5] ISO/IEC JTC1/SC29/ WG11, Doc. MPEG2003/N6193, Call For Proposals on Scalable Video Coding Technology, Waikoloa, December 2003.

[6] K. Ugur, P. Nasiopoulos, Design issues and proposal for H.264 based FGS, ISO/IEC JTC 1/SC 29/WG 11 Doc. MPEG2003/M9505, Pattaya, March 2003.

[7] 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.

[8] B. Shen, I. K. Sethi, and V. Bhaskaran, "Adaptive motion vector resampling for compressed video downscaling," *Proc. ICIP'97*, pp. 771 - 774, October 1997.

[9] P. Yin, M. Wu, and B. Liu, "Video transcoding by reducing spatial resolution," *Proc. ICIP'00*, pp. 972 - 975, September 2000.

[10] A. Vetro, Ch. Christopoulos, H. Sun, Video transcoding and architectures and techniques: an overview, *IEEE Signal Proc. Magazine*, vol. 20, pp. 18-29, March 2003.

[11] Lange R., Domański M., "Motion vector encoding in AVC spatially-scalable video coders", *Proc. 10th Intern. Workshop Systems, Signals and Image Proc., IWSSIP'2003*, Prague, Sept. 10-11, 2003, pp. 111-114.