EXTENDED INTER-VIEW DIRECT MODE FOR MULTIVIEW VIDEO CODING

Jacek Konieczny, Marek Domański

Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology, Poznań, Poland

ABSTRACT

This paper presents a novel compression tool that improves compression performance of Multiview Video Coding when depth information is available for the reference view. This tool is proposed as a new Extended Inter-View Direct mode of macroblock coding and exploits improved inter-view prediction of motion information. For this prediction, the idea is to use depth information in order to predict motion vectors more accurately, as it is possible in the classic Multiview Video Coding. The Extended Inter-View Direct mode is designed especially for the purpose of joint video and depth coding, where depth information for coded view is not available. Extensive experiments show the potential of the new mode to reduce bitstream by 2% to 11% depending on test sequence, compression scenario and variant of the state-of-the-art multiview compression reference technique used.

Index Terms— Inter-View Direct mode, motion vector prediction, Multiview Video Coding, depth-enhanced video.

1. INTRODUCTION

Recently, MPEG-4 AVC/H.264 standard has been extended by Multiview Video Coding (MVC) [1] that exploits mutual correlation between views in order to reduce total bitrate of the whole multiview video. Although MVC has been already adopted in some applications, its compression efficiency seems not to be satisfactory for future 3D applications. Therefore a great research effort is being made towards new compression techniques for multiview and 3D video. This paper remains within this mainstream. The main idea of this paper is to exploit additional depth information in order to improve the overall compression performance of MVC.

Encoded video bitstream consists mainly of three types of data, i.e. control data, texture residuals and motion information. Here, a new technique for efficient representation of the motion data is proposed that results from the observation that motion information in neighboring views in a multiview video sequence is often similar. Thus, reduction of this information redundancy by reusing motion vectors from reference views should likely improve the achieved compression ratio.

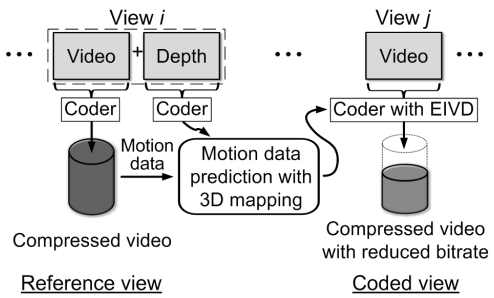

Fig. 1. Motion data prediction with 3D mapping.

Based on a similar concept, an additional compression mode called Motion Skip (MS) has been already proposed for MVC in [2,3]. In this mode, a global disparity vector is used for each encoded macroblock to identify the initial position of corresponding motion vectors in reference frames. Next, representative motion vectors are selected during an additional search operation and then reused.

In contrast, this paper describes a more advanced technique for prediction of motion data in multiview video. This new technique exploits more accurate description of 3D video scene but is much less computationally complex.

2. PROPOSED APPROACH

Here, motion information for one view is predicted from depth and motion information available for reference views using a 3D mapping (Fig. 1). Consequently, motion information encoded in the bitstream can be reduced. This results in improved compression efficiency, especially for lower bitrates.

The preliminary concept was introduced in [4] as a new macroblock compression mode called the Inter-View Direct (IVD) mode. In the IVD mode, depth information is used to define a mapping between each pixel in the encoded frame and its counterpart pixel in the reference frame. Using this mapping a motion vector can be obtained for each pixel in the encoded image by a simple derivation of the motion information assigned to the corresponding pixel in the reference view, except of the intra-coded pixels.

However, previously proposed IVD mode [4] requires depth information for both encoded and reference views to establish the inter-view mapping between pixels. Unfortunately, this makes it non-applicable while joint video and depth coding or depth estimation at the decoder is performed. To handle these important issues, in this paper we propose a new Extended Inter-View Direct (EIVD) mode. In this mode, depth information for the encoded view is not used at the time of encoding or decoding. As a result, the new mode requires depth information associated only with the reference views, which solves the problem of adopting the IVD mode to most of the multiview video applications.

3. EIVD MODE

The proposed EIVD mode that can be selected for a macroblock during rate-distortion optimization in the encoder. In EIVD mode,

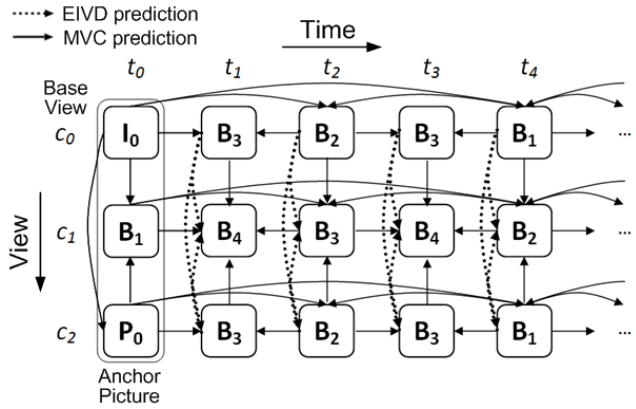

Fig. 2. Prediction scheme using EIVD mode for multiview video coding.

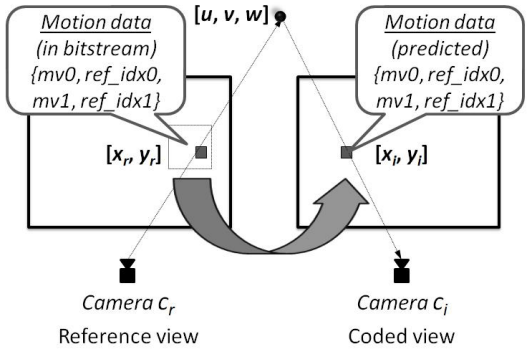

Fig. 3. Multiview point correspondence.

motion information for a macroblock, i.e. motion vectors and reference indices, is inferred directly from the macroblocks already encoded in the neighboring view which was selected as a reference view. Consequently, no motion information is transmitted in the bitstream, as the same procedure is carried out in the decoder.

Let us now refer to the prediction scheme used in MVC (see Fig. 2). Here, three views are encoded in the following order: c_0, c_2 and c_1.The EIVD mode may be used for both c_1 and c_2 views. In case of view c_2, the EIVD mode can be applied using reference frames from the view c_0. Similarly, view c_1 can be encoded utilizing the EIVD mode with reference frames from views c_0 and c_2. However, when the base view c_0 is encoded there are no reference frames from other views encoded already and the EIVD mode cannot be applied. Additionally, EIVD mode is also disabled in the case of anchor pictures as defined in JMVM [5], because no motion information is available in the reference views (Fig. 2).

Let us denote the view related to the currently encoded picture as c_i and the reference view as c_r. The algorithm extends the one presented in [4] with a modification which makes the encoding process independent of the depth map (or disparity map) of the encoded view c_i. Consequently, the whole procedure of motion data prediction may be carried out without information about depth (disparity) of the view c_i, which enables application of the algorithm in case of joint video and depth coding. The main steps of the algorithm are listed as follows (refer to Fig. 3):

1) *Find corresponding pixel locations*. For each pixel of the reference view picture (view c_r) pixel location (x_r,y_r) is projected from view c_r into 3D world coordinates (u,v,w) using Depth-Image Based Rendering technique (DIBR) and re-projected into local coordinates (x_i,y_i) of the coded view picture c_i. In this process, depth values for all pixels of reference view pictures must be provided, e.g. in the form of a disparity map. DIBR also requires camera parameters, which are usually presented in the form of intrinsic and extrinsic camera matrices for each view.
2) *Occlusion test*. Corresponding pixel positions (x_i,y_i) and (x_r,y_r) are tested for occlusion. If a pixel (x_i,y_i) has no corresponding pixel (x_r,y_r) assigned to it after projecting all the pixel locations from the reference view c_r - it is assumed to be occluded.
3) *Derive motion information*. If a coded pixel (x_i,y_i) is not occluded and the corresponding pixel (x_r,y_r) provides all required information for inter-frame prediction mode, pixel (x_r,y_r) is marked as a source of motion information for the coded pixel (x_i,y_i).
4) *Motion compensation*. After determining the source of motion information for each point (x_i,y_i) of the coded macroblock, the macroblock is motion compensated independently for each point using motion vectors and reference indices derived from the source point (x_r,y_r).

In cases where motion information is not available for coded pixel (x_i,y_i) after steps *1*, *2* and *3*, e.g. when pixel (x_i,y_i) is occluded or a corresponding pixel (x_r,y_r) was intra-predicted, the algorithm tests some other possibilities to find the corresponding pixel (x_r,y_r). Firstly, other available reference views are inspected according to the inversed view coding order. If no reference view is available, motion information for pixel (x_i,y_i) is derived from the neighboring points of the same view, assuming that these points have corresponding pixels in the reference. Finally, if this cannot be applied, the standard Direct mode prediction is used to gain the motion information for pixel (x_i,y_i).

In the implementation of the algorithm it is recommended to use a dedicated buffer to store the corresponding pixel locations (x_r,y_r) for the whole encoded picture. This preserves encoder and decoder from re-projecting the pixel locations for each macroblock. As a result, the only computational overhead in the encoder and decoder is related to the once-per-frame depth-image based rendering (DIBR) and occluded point procedures, which influence the processing time quite slightly. Consequently, the proposed method increases the encoder and decoder complexity negligibly in comparison with the state-of-the-art MVC. Additionally, when compared with the IVD mode [4], the proposed EIVD mode reduces computational cost of inter-view motion information prediction as it requires projecting the pixel positions from reference view into coded view only.

4. SYNTAX OF EIVD MODE

The Extended Inter-View Direct mode was integrated into MVC reference software model [1] with appropriate changes to the multiview bitstream syntax and semantics.

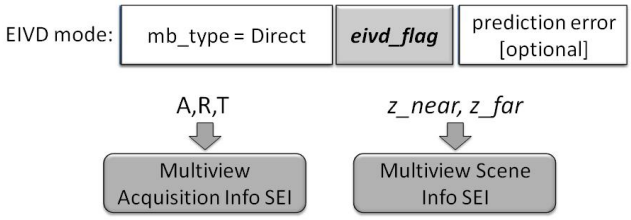

Fig. 4. Syntax modifications.

Table 1. Compression performance of JMVC+EIVD compared to compression performance of JMVC and JMVM+MS for *Scenario1* and *Scenario 2* using Bjontegaard measures.

QP 24,30,36,42	*Scenario 1*				*Scenario 2*			
	JMVC		*JMVM+MS*		*JMVC*		*JMVM+MS*	
	ΔPSNR[dB]	*ΔBitrate[%]*	*ΔPSNR[dB]*	*ΔBitrate[%]*	*ΔPSNR[dB]*	*ΔBitrate[%]*	*ΔPSNR[dB]*	*ΔBitrate[%]*
Book Arrival	0.33	-9.2	0.47	-12.9	0.22	-6.0	0.10	-2.7
Newspaper	0.22	-5.2	0.54	-12.3	0.17	-4.0	0.26	-6.1
Lovebird1	0.15	-4.6	0.47	-13.6	0.09	-2.9	0.34	-10.3
Champagne tower	0.32	-7.2	0.40	-9.0	0.26	-5.9	0.01	-0.4
Pantomime	0.40	-8.3	0.42	-8.6	0.35	-7.7	-0.23	5.4
Average	0.28	-6.9	0.46	-11.3	0.22	-5.3	0.10	-2.8

Usage of EIVD mode is signaled to the decoder with a new flag included in the bitstream: *eivd_flag*, a modification to the existing Direct mode macroblock layer syntax in non-anchor pictures of non-base views only. An additional bit representing the *eivd_flag* is added if *mb_type* is signaling the Direct mode selection to distinguish the new EIVD mode from the traditional Direct mode. As a consequence, the base view syntax is not modified and remains fully compliant with MPEG-4 AVC/H.264.

Similarly as proposed in [4], the order of reference views to be selected for the motion data derivation in EIVD mode is based on the view coding order included originally in the MVC bitstream.

The EIVD mode requires that intrinsic and extrinsic camera parameters matrices (A,R,T) are transmitted (Fig. 4). Additionally, in order to use disparity maps as the depth information source, the *z_near* and *z_far* values are needed in the decoder. This information may be included in SEI messages as proposed originally in [4,6,7].

5. EXPERIMENTAL RESULTS

The goal of the experiments was to assess the impact of the new EIVD mode onto the multiview video compression performance.

5.1. Experiment setup

The EIVD mode was implemented into the JMVC 4.0 software (JMVC+EIVD) [8]. Such a codec was compared with the codecs:
1) original JMVC 4.0 software (JMVC) [9],
2) JMVM 8.0 software with enabled motion skip and single loop filtering (JMVM+MS) [10].

Two coding scenarios of stereo sequence encoding were investigated. *Scenario1*: encode view c_2 (Fig. 2) with the c_0 view as the reference view – only anchor reference pictures of one reference view are available. *Scenario2*: encode view c_1 (Fig. 2), with only c_0 available as the reference view. In this scenario both anchor and non-anchor reference pictures are used.

Five standard multiview test sequences were used: "Book Arrival" [11], "Newspaper" [12], "Lovebird1" [13], "Champagne tower" [14] and "Pantomime" [14]. The sequences were encoded with hierarchical B frames, GOP size equal 12 and quantization parameter QP = {24, 30, 36, 42}. Quality of decoded video was measured by luma PSNR (PSNRY) averaged over the first 96 frames of each test sequence.

5.2. Results

For both scenarios, Figs. 5 and 6, and Table 1 present improvement in compression performance obtained for the codec with Extended Inter-View Direct mode (JMVC+EIVD) against the standard MVC codec (JMVC) and codec with Motion Skip mode (JMVC+MS) using the Bjontegaard measures [15]. As the proposed technique is designed for applications that use depth information regardless of the method of delivering it, presented bitrates are calculated for an encoded view only. Bits needed for transmission of depth information are not included. We assume that depth information is already transmitted for other purposes.

The proposed EIVD mode always improves compression efficiency of multiview video codec. The average bitrate saving is about 6.9% for *Scenario1* and about 5.3% for *Scenario2* (see comparison JMVC+EIVD against JMVC in Table 1 and 2). Also, on average, EIVD mode improves multiview compression performance better than Motion-Skip. When compared to JMVM+MS, the average bitrate savings of JMVC+EIVD are about 11.3% for *Scenario1* and 2.8 for *Scenario2*. In *Scenario1* the gain is larger for JMVM+MS than for JMVC because Motion-Skip adds extra bit to each mackroblock header despite it is used only in anchor pictures. However, for some sequences in *Scenario2*, the compression efficiency of JMVM+MS is better than for

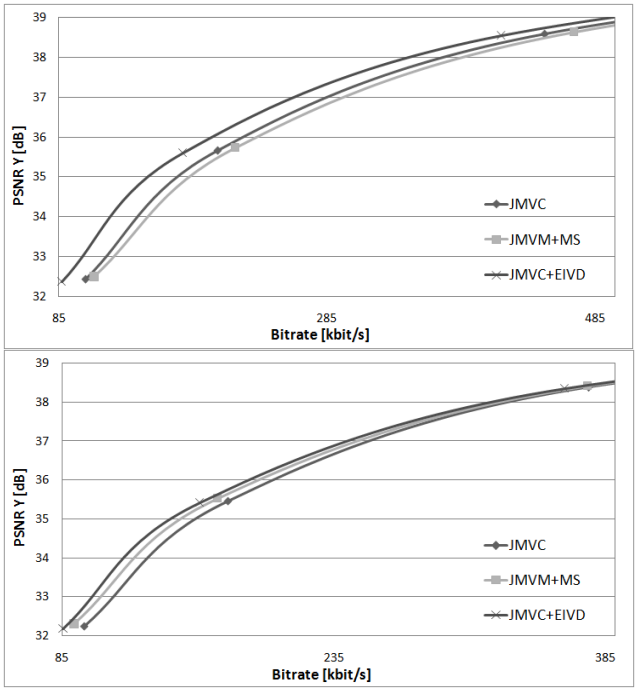

Fig. 5. Rate-distortion curves: top – *Scenario1*, bottom – *Scenario2*, sequence: "Book Arrival".

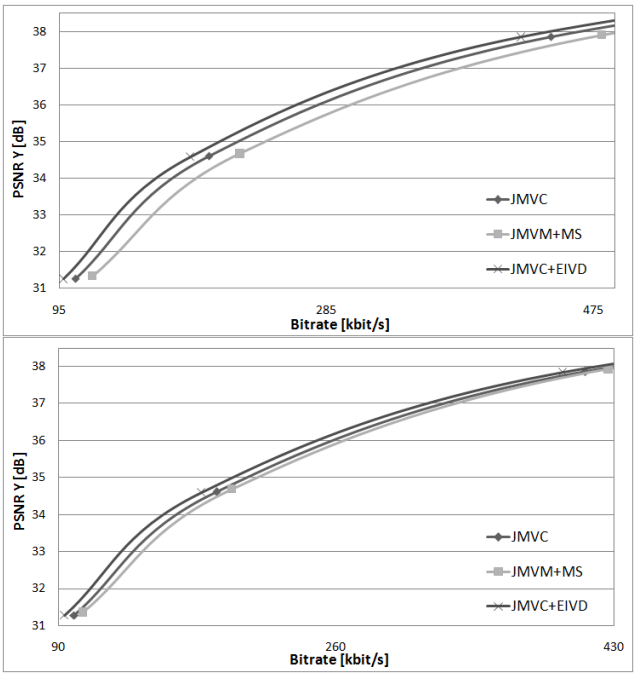

Fig. 6. Rate-distortion curves: top – *Scenario1*, bottom – *Scenario2*, sequence: "Newspaper".

JMVC+EIVD. This is noticeable especially for the "Pantomime" sequence. It should also be noticed that the overall compression efficiency for *Scenario1* is worse than for *Scenario2* for all tested cases due to unused inter-view correlation in the original MVC scheme (Fig .2).

In the experiment the new EIVD mode was also compared to the previous IVD mode [4] in order to explore the impact of usage of only the reference view depth map on compression performance. Experimental results show that this impact is very slight (average bitrate reduction of 0.5% for *Scenario1* and 0.4% for *Scenario2* against IVD). However, the new EIVD mode needs roughly 50% of computational power used by the IVD mode.

As the EIVD mode utilizes depth information in the encoding process, compression performance depends on accuracy of depth. Presented results were obtained for available standard depth maps. Probably, limited quality of depth maps used for "Pantomime" sequence is the reason of reduced efficiency of EIVD mode for this sequence.

6. CONCLUSION

In this paper, a new macroblock compression mode for inter-view prediction of motion vectors for efficient encoding of multiview video was introduced. This new mode is an extension of previously proposed Inter-View Direct mode.

Extended Inter-View Direct mode is designed for depth-enriched multiview video applications, as it requires only reference view depth information. As a consequence, the proposed mode is suitable especially for joint video and depth coding scenarios. At the same time, the proposed Extended Inter-View Direct mode does not reduce compression efficiency of the original Inter-View Direct mode and also substantially reduces computational overhead of this algorithm.

Experiments show that the proposed mode clearly improves compression efficiency of the state-of-the-art multiview video coding. Achieved bitrate savings are usually between 2% and 11% with negligible impact on the encoder and decoder complexity. However, further increase in accuracy of depth maps will probably raise the importance of this method.

The conclusion is that the proposed extension of motion data inter-view prediction technique exhibits some potential in compression of depth–enriched multiview video that is likely to be an important issue in 3D video technology.

7. ACKNOWLEDGEMENTS

This work was supported by the public funds as a research project.

8. REFERENCES

[1] ISO/IEC 14496-10 (MPEG-4 AVC) / ITU-T Rec. H.264: Advanced Video Coding for Generic Audiovisual Services (2009).

[2] H.-S. Koo, Y.-J. Jeon, B.-M. Jeon, "Motion Skip Mode for MVC", ITU-T and ISO/IEC JTC1, JVT-U091, Hangzhou, China, October 2006.

[3] H. Yang, Y. Chang, J. Huo, "Fine-Granular Motion Matching for Inter-View Motion Skip Mode in Multiview Video Coding", Circuits and Systems for Video Technology, IEEE Transactions on, vol. 19, Issue 6, pp. 887–892, June 2009.

[4] J. Konieczny, M. Domański, "Depth-Based Inter-View Prediction of Motion Vectors for Improved Multiview Video Coding", 3DTV-CON 2010, Tampere, Finland, June 2010.

[5] A. Vetro, P. Pandit, H. Kimata, A. Smolic, "Joint Draft 2.0 on Multiview Video Coding", ITU-T and ISO/IEC JTC1, JVT-V209, Marrakech, Morocco, January 2007.

[6] A. Vetro, S. Yea, W. Matusik, H. Pfister, M. Zwicker, "Antialiasing for 3D Displays", ITU-T and ISO/IEC JTC1, JVT-W060, San Jose, USA, April 2007.

[7] S. Yea, A. Vetro, M. Zwicker, "MVC Showcase for SEI on Scene and Acquisition Info", ITU-T and ISO/IEC JTC1, JVT-X074, Geneva, Switzerland, June 2007.

[8] J. Konieczny, JMVC 4.0 + EIVD mode software, http://www.multimedia.edu.pl/publications/software/index.htm.

[9] Y. Chen, P. Pandit, S. Yea, "WD 4 Reference software for MVC", ITU-T and ISO/IEC JTC1, JVT-AD207, Geneva, Switzerland, February 2009.

[10] P. Pandit, A. Vetro, Y. Chen, "JMVM 8 software", ITU-T and ISO/IEC JTC1, JVT-AA208, Geneva, Switzerland, April 2008.

[11] I. Feldmann, M. Mueller, F. Zilly, R. Tanger, K. Mueller, A. Smolic, P. Kauff, T. Wiegand, "HHI Test Material for 3D Video", ISO/IEC JTC1/SC29/WG11 MPEG 2008/M15413, Archamps, France, April 2008.

[12] Y.-S. Ho, E.-K. Lee, C. Lee, "Multiview Video Test Sequence and Camera Parameters", ISO/IEC JTC1/SC29/WG11 MPEG2008/M15419, Archamps, France, April 2008.

[13] G.-M. Um, G. Bang, N. Hur, J. Kim, Y.-S. Ho, "3D Video Test Material of Outdoor Scene", ISO/IEC JTC1/SC29/WG11 MPEG2008/M15371, Archamps, France, April 2008.

[14] M. Tanimoto, T. Fujii, N. Fukushima, "1D Parallel Test Sequences for MPEG-FTV", ISO/IEC JTC1/SC29/WG11 MPEG2008/M15378, Archamps, France, April 2008.

[15] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves", VCEG Contribution VCEG-M33, Austin, USA, April 2001.