Depth-Enhanced Compression for 3D Video

Jacek Konieczny, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology
Poznań, Poland

***Abstract* — The original idea of the paper is to exploit depth maps in order to increase compression efficiency for multiview video. Depth information is used to establish a 3D mapping between each pixel in an encoded frame and its counterpart in the reference view. With this mapping motion vectors and reference frame indices can be obtained independently for each pixel in a coded picture by a simple derivation of the motion information assigned to the corresponding pixel in the reference view. The goal of the paper is to explore a practical way to use this idea in multiview video coding. This goal has been achieved by experimental testing of various sets of direct and skip modes with and without inter-frame motion prediction. It was shown that either "all modes" have to be used or "all but not classic direct mode" should be used. The respective experimental results have been provided in the paper.**

I. INTRODUCTION

A new challenge in video compression is related to prospective new generation of 3D video systems. Among many applications of such systems, the second generation 3D television is one of the most promising. Such systems comprise video broadcasting to autostereoscopic displays as well as free-viewpoint television (FTV). For the abovementioned applications, video taken from many viewpoints must be available for display at a receiver. Currently, the number of views for good-quality autostereoscopic displays is about 30. The number of views in FTV systems may be even higher. Obviously, simulcast transmission of a video that corresponds to such a large number of simultaneous views would not be practical.

The state-of-the-art standard MPEG-4 AVC/H.264 [1] provides Multiview Video Coding (MVC) extension that standardizes video compression exploiting the inter-view redundancy. As the bitrate reduction is usually only about 10-30% of the bitrate for simulcast, this technique is not sufficient for the abovementioned applications. Therefore, in the year 2010, MPEG expert group of ISO/IEC has announced "Call for Proposals on 3D Video Coding Technology" [2]. In this Call for Proposals (CfP), MVC is mentioned as the reference for further research and comparisons. Also in this paper, some extensions of the MVC technology are described.

For the new generation of 3D video systems, variants of "multiview video plus depth" are considered as potential transmission formats. This paper also continues such an approach. The original idea is to exploit depth maps in order to increase compression efficiency for multiview video. This approach is unlike in many other papers where depth representation is proposed to be improved by the use of the information extracted from the respective video.

In this paper, we assume that video and depth are transmitted for a base view. A dependent view is encoded and decoded with reference to the available information from the base view that is also called the reference view. For texture encoding, inter-view prediction may be used as it has been already extensively studied for MVC. However, in this paper we will focus on the usage of motion vectors inherited from the reference view (Fig. 1).

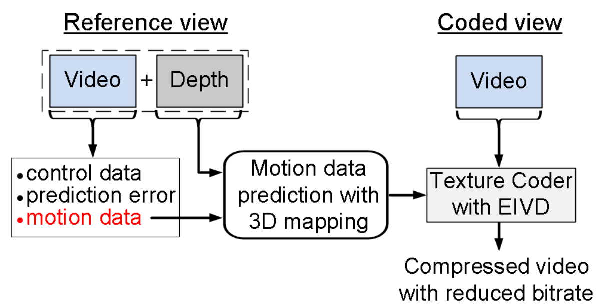

Figure 1. Motion data prediction with 3D mapping.

II. THE DBMP ALGORITHM FOR MOTION PREDICTION

The idea of Depth-Based Motion Prediction (DBMP) is based on the observation that motion information (i.e. motion vectors) for one view can be predicted from depth and motion information available for a reference view. As a result, application of the abovementioned coding tool yields reduced bitrate for encoding of motion information. And consequently, the overall compression efficiency of the codec improves, especially for lower bitrates. The concept of the DBMP algorithm was initially introduced for a new macroblock mode called the Extended Inter-View Direct mode [3,4,5]. In this paper, this name is abbreviated to Inter-View Direct (IVD) mode. The goal of this paper is to study how Depth-Based Motion Prediction (DBMP) and Inter-View Direct (IVD) mode may be efficiently used in order to increase compression efficiency for multiview video enhanced by depth.

For the sake of brevity, we are going to skip detailed description of DBMP and the IVD mode that may be found in [5]. In this mode, depth information is used to establish a 3D mapping between each pixel in the encoded frame and its counterpart in the reference view. With this mapping motion vectors and reference frame indices can be obtained independently for each pixel in the encoded picture by a simple derivation of the motion information assigned to the corresponding pixel in the reference view (Fig. 1). This new depth-based mapping of motion vectors is more sophisticated and more exact than the 2D mapping used in MVC.

In the IVD mode, motion information for each pixel in this macroblock, i.e. motion vectors and reference picture indices, is directly inferred from the macroblocks already encoded in the reference view. As the same procedure is repeated in the decoder, no motion information is transmitted in the bitstream. This yields that IVD mode is a very efficient tool for encoding of motion information. However, it cannot be applied to the first view in the coding order, which is called the base view. Similarly, the IVD mode is disabled in the case of the anchor pictures as defined in [1], because no motion information referring to other time instances is available in the reference views.

Depth information for the encoded view is not required. Therefore, Depth-Based Motion Prediction (DBMP) and IVD mode are applicable for most multiview applications, including joint video and depth coding.

III. Mode Selection Strategy

The Depth-Based Motion Prediction (DBMP) may be used not only for the IVD mode but also for the Inter View Skip (IVS) mode where no texture prediction error is transmitted. So, we have four Direct modes for macroblock coding:

- standard Skip
- standard Direct,
- new IVS mode (Inter View Skip),
- new IVD mode (Inter View Direct).

Therefore, the question arises if all the modes are necessary. The papers [3-5] explain that the IVD (called also EIVD) mode yields an increase of compression efficiency. Then, there is also the question of what the right way of mode signaling in a bitstream is. We are going to answer this question here. We are going to present a complex study on possible mode selection strategies and show necessary syntax modifications to implement the codec.

The IVD mode introduced in [3,4,5] was originally signaled to a decoder with a new flag included in the bitstream: *eivd_flag*. The idea was to modify the standard Direct mode macroblock layer syntax in non-anchor pictures of non-base views only. An additional bit representing *eivd_flag* is added if *mb_type* is signaling the Direct mode selection to distinguish the new IVD mode from the traditional Direct mode.

In general, we propose to use four syntax elements to carry the information related to Direct modes:

- *skip_flag* that signals Skip mode,
- *mb_type* – type of a given macroblock,
- *eivd_flag* – 1-bit flag related to IVD mode,
- *cbp* that communicates the pattern of encoded and uncoded blocks (0 means that all blocks are empty).

Let us now refer to Fig. 2. Here, eight different variants of syntax to signal Direct and Skip modes are presented. For some variants less than four Direct modes are available.

The standard AVC codec uses the syntax *variant 1*: Skip mode is signaled with a *skip_flag*, in other cases the Direct mode is signaled with a full macroblock header including *mb_type* specific for the Direct mode, IVD mode is not used. In principle, the signaled variant for the IVD mode is *variant 2*. The Direct mode has the *eivd_flag*=0 added to distinguish it from the IVD mode, IVD and IVS modes are discriminated by non-zero prediction error - we use coded block pattern *cbp* for that purpose.

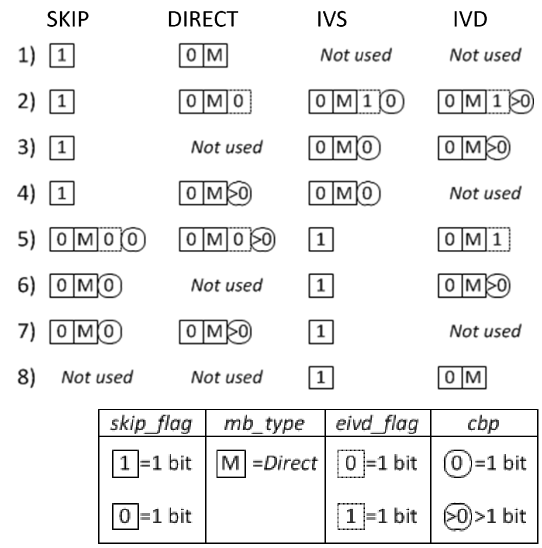

Figure 2. Codec variants and the respective syntax structures.

Our first observation is that Direct and IVD modes (with non-zero prediction error) are selected rarely, especially for lower bitrates. Disabling one of these modes will increase the prediction error for some macroblocks, but coding gain may be achieved by not sending the additional 1-bit flag *eivd_flag*. For that purpose codec *variants 3* and *4* were designed (compare with *variant 2*). In *variant 3* the Direct mode is not used. Consequently, *eivd_flag* is not necessary to distinguish it from IVD modes. In *variant 4* the IVD mode is not used. IVS is distinguished from Direct by testing if *cbp* is set to zero, so *eivd_flag* is also redundant.

Another issue is that the motion prediction algorithm used in IVD may be more efficient than the traditional median prediction of motion vectors and selecting first pictures on the list as the reference. As we observed, the IVD algorithm generally gives a prediction error similar to the median prediction; however, especially in cases of complex motion, it may perform better. In order to verify this, IVS was signaled with the syntax of Skip mode, resulting in codec variants *5-7* (corresponding to variants *2-4*). Here, IVS is signaled with *skip_flag*=1 and other modes are encoded with a full macroblock header including *mb_type* specific for Direct. The last codec variant, *variant 8*, was implemented to check if the traditional Direct mode could be substituted by the IVD mode.

IV. Experimental Results And Analysis

The goal of the experiment was to assess the impact of the new mode selection strategies onto the multiview video compression performance.

A. Experiment setup

The IVD and IVS modes were implemented into the JMVC 4.0 software (codec variants *2-8*) [6] and compared with the original JMVC 4.0 software (codec variant *1*) [7].

Two coding scenarios of stereo sequence encoding were investigated. *Scenario1*: encoding with one reference view and only anchor reference pictures for inter-view prediction. *Scenario2*: one reference view and both anchor and non-anchor reference pictures for inter-view prediction.

Five standard multiview test sequences were used: "Book Arrival" [8], "Newspaper" [9], "Lovebird1" [10], "Champagne tower" [11] and "Pantomime" [11]. The sequences were encoded using CABAC entropy coder, with hierarchical B frames, GOP size equal 12 and quantization parameter QP = {27, 30, 33, 36}. Quality of the decoded video was measured by luma PSNR (PSNR-Y) averaged over the first 96 frames from each sequence.

TABLE I. COMPRESSION PERFORMANCE OF CODEC VARIANTS 2-8 COMPARED TO VARIANT 1 FOR SCENARIO 1 AND SCENARIO 2 USING BJONTEGAARD MEASURES.

QP 27, 30, 33, 36	**dPSNR-Y [dB]**							**dBitrate [%]**						
	***Scenario 1* (without non-anchor reference)**													
Sequence\Codec variant	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Book Arrival	0.20	0.20	0.17	0.24	0.23	0.19	0.22	-5.7	-5.7	-5.0	-6.8	-6.5	-5.4	-6.3
Champagnetower	0.29	0.30	0.23	0.35	0.36	0.28	0.33	-6.5	-6.9	-5.2	-8.0	-8.0	-6.5	-7.6
Lovebird1	0.13	0.13	0.10	0.13	0.14	0.10	0.14	-4.0	-4.0	-3.0	-4.1	-4.2	-3.0	-4.3
Newspaper	0.20	0.20	0.18	0.23	0.21	0.19	0.17	-4.9	-5.0	-4.6	-5.6	-5.3	-4.7	-4.4
Pantomime	0.40	0.40	0.30	0.44	0.44	0.34	0.40	-8.6	-8.7	-6.5	-9.5	-9.5	-7.4	-8.7
Average	0.24	0.25	0.20	0.28	0.28	0.22	0.25	-5.9	-6.1	-4.9	-6.8	-6.7	-5.4	-6.2
	***Scneario 2* (with non-anchor reference)**													
Sequence\Codec variant	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Book Arrival	0.12	0.13	0.12	0.16	0.15	0.14	0.12	-3.6	-3.8	-3.6	-4.7	-4.4	-4.0	-3.4
Champagnetower	0.25	0.26	0.21	0.31	0.31	0.26	0.28	-5.6	-5.9	-4.8	-7.2	-7.1	-6.1	-6.5
Lovebird1	0.08	0.07	0.06	0.07	0.08	0.06	0.07	-2.5	-2.3	-2.1	-2.4	-2.4	-1.9	-2.4
Newspaper	0.14	0.15	0.15	0.17	0.15	0.15	0.11	-3.6	-3.8	-3.7	-4.3	-3.9	-3.7	-2.9
Pantomime	0.33	0.33	0.30	0.40	0.38	0.35	0.32	-7.7	-7.8	-7.0	-9.3	-9.0	-8.3	-7.5
Average	0.19	0.19	0.17	0.22	0.21	0.19	0.18	-4.6	-4.7	-4.2	-5.6	-5.3	-4.8	-4.5

B. Results

Fig. 3 and Table I present improvement in compression performance obtained for codec variants *2-8* (codecs with IVD/IVS modes implemented) against variant *1* using the Bjontegaard measures [12]. These measures reflect average changes of PSNR-Y and bitrate of the encoded dependent view. The depth information is assumed to be transmitted with the base layer for other purposes.

The results show that adding the IVD prediction increases the frequency of low-cost macroblock modes (Direct, Skip, IVD, IVS) selection regardless of the selected mode selection strategy (variants *2-8*). As a result, macroblock modes that require transmitting motion vectors are sparsely selected (difference is 1.7-5.3 [pp] on average) and thus, the compression performance is better. The achieved bitrate reductions are equal 4.9-6.8 [%] in the case of *Scenario 1* and 4.2-5.6 [%] in the case of *Scenario 2* (Table I). We also see that coding gains are bigger for lower bitrates.

We may observe that modifications of the mode selection strategy changed the compression performance of the codec. The average bitrate reduction, compared to the original IVD syntax (variant *2*), was increased by up to 1.0 [pp] (variant *5*). However, not all codec variants performed better than variant *2*. More significant changes can be distinguished in frequency of selection of different macroblock modes. Fig. 4 shows the average macroblock mode selection of Direct, Skip, IVD and IVS modes for all test sequences. Fig. 5 presents the total mode selection of Direct and Skip, IVD and IVS and the four abovementioned modes.

A comparison between codec variants *1-4* and *5-8* shows that IVD prediction of motion information performs better than the traditional prediction used in the Direct mode as far as bitrate reduction is concerned. The main reason is the frequency of low-cost modes selection. If we consider the "Sum" columns in Fig. 5 we can see they are always higher for variants *5-8* than for their counterpart variants *1-4*.

Let us refer to the issue of disabling Direct or IVD modes in the codec (variants *3*,*4*,*6* and *7*). In the case of variant *3* a slight bitrate reduction is observed in Fig. 3 (0.1-0.2 [pp]) against variant *2* which comes from more frequent usage of macroblocks with DBM prediction (1.5 [pp] gain) due to lack of *eivd_flag* to signal IVD modes. In the variant *4* however, we noted loss of coding performance when compared to variant *2*.

In this case, due to lack of efficient way of encoding complex motion in low-cost form of the IVD mode, some of macroblocks were encoded with extra motion information resulting in higher bitrate. In the case of variants *6* and *7* (compared to variant *5*) the gain from not transmitting additional *eivd_flag* is lower than the loss on larger prediction error. However, it is also better to disable the Direct mode instead of IVD in these cases. To conclude, concerning the results it seems that the best mode selection strategy is preserving all the macroblock modes to match various cases, while setting the IVS as the least expensive mode (variant *5*).

Another observation comes from the comparison of variant *8* performance for *Scenario 1* and *2* (Fig. 3). In both cases we note bitrate reduction against codec without IVD prediction (variant *1*). However, lack of Direct and Skip modes in *Scenario 2* results in worse codec performance when compared to *Scenario 1*. The reason is that a multiview codec uses Direct and Skip modes as inter-view prediction methods if non-anchor references are allowed (*Scenario 2*).

Consequently, disabling these modes in *Scenario 2* results in bigger loss of coding performance than for *Scenario 1* where these modes are unavailable for inter-view prediction.

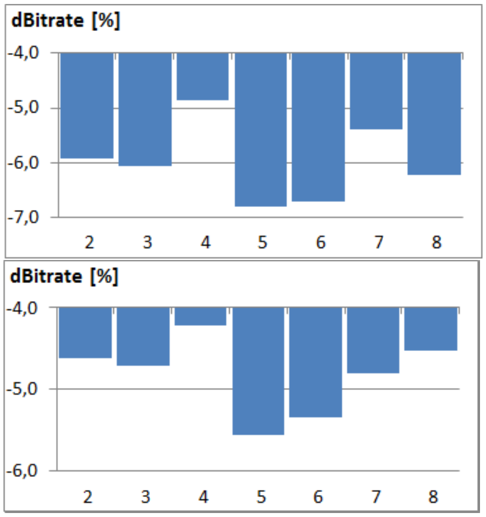

Figure 3. Bjontegaard measures for codec variants 2-8 vs. variant 1 (average over all test sequences): top - *Scenario 1*, bottom – *Scenario 2*.

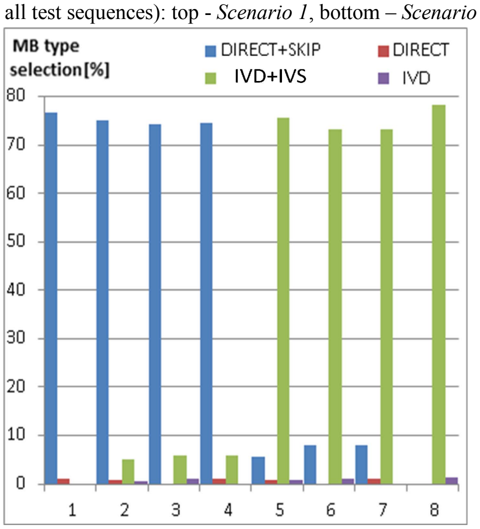

Figure 4. Average macroblock mode selection for codec variants 1-8.

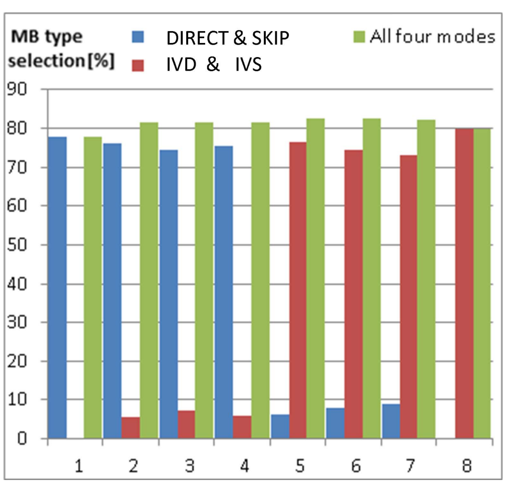

Figure 5. Average macroblock mode selection for Direct, IVD and both motion data predictions for codec variants 1-8.

V. Conclusion

In this paper, a new compression strategy for depth-enhanced multiview sequences has been proposed. The strategy exploits the idea of new macroblock compression modes called Inter-View Direct and Inter-View Skip. Experiments show that adding the proposed Inter-View Direct (IVD) mode to the low-cost modes set of the codec decreases the number of macroblocks for which transmitting an additional motion information is required. Consequently, the coding performance increases, resulting in bitstream reduction up to 6.8 [%] on average. However, the relative coding gains grow for lower bitrates, as the cost of encoding motion information is higher for small bitstreams.

Experimental results show also that, despite introducing the new IVD mode, all the existing macroblock modes should be preserved in the codec to assure the best performance in various cases. In average, removal of classic Direct mode from the mode list would result in negligible loss of compression performance (see variant 6).

To conclude, the proposed IVD and IVS modes implement new motion data inter-view prediction technique which may be successfully adopted into developed AVC-compatible codecs for depth-enriched multiview video.

References

[1] ISO/IEC 14496-10 (MPEG-4 AVC) / ITU-T Rec. H.264: Advanced Video Coding for Generic Audiovisual Services (2010).

[2] “Call for Proposals on 3D Video Coding Technology”, ISO/IEC JTC1/SC29/WG11, MPEG2011/N12036, Geneva, Switzerland, March 2011.

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

[4] J. Konieczny, M. Domański, “Inter-View Direct Mode for Multiview Video Coding”, ISO/IEC JTC1/SC29/WG11, MPEG2010/M17800, Geneva, Switzerland, July 2010.

[5] J. Konieczny, M. Domański, “Extended Inter-View Direct Mode for Multiview Video Coding”, ICASSP 2011, Prague, Czech Republic, May 2011.

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

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

[8] 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.

[9] 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.

[10] 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.

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

[12] G. Bjontegaard, “Calculation of Average PSNR Differences between RD-curves”, VCEG Contr. VCEG-M33, Austin, USA, April 2001.