Depth-Based Inter-View Motion Data Prediction for HEVC-Based Multiview Video Coding

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

Abstract—**The paper deals with efficient coding of motion data for compression of multiview video with depth maps. This research was done in the context of the new compression technology called High Efficiency Video Coding. In the proposed approach, motion vectors and reference frame indices are predicted from reference view for each pixel using the Depth-Based Motion Prediction (DBMP). The new codec with DBMP compression tool was compared to the multiview HEVC-based codec i.e. a set of HEVC codecs augmented with inter-view texture prediction tools adopted from classic multiview codec (MVC) based on the AVC technology. Extensive experiments show the potential of the new DBMP predictor to reduce bitrate up to 12% against the multiview HEVC-based codec.**

Keywords: Depth-Based Motion Prediction, Inter-View Direct mode, multiview video coding, depth-enhanced video, HEVC.

I. INTRODUCTION

The ultra-high resolution as well as 3D video are the new challenges related to the video compression technology.

Among many applications of the 3D video systems, one of the most promising is the second generation 3D television with video broadcasting to autostereoscopic displays as well as free-viewpoint television (FTV). For such applications, multiple viewpoints of recorded scene must be available at a receiver. Currently, the number of views for good-quality autostereoscopic displays approach 30. In case of FTV systems this number may be even higher. Simulcast transmission of such a large number views is obviously unpractical. The state-of-the-art MPEG-4 AVC/H.264 standard [1] provides Multiview Video Coding (MVC) extension that describes video compression exploiting the inter-view redundancy. Unfortunately, the bitrate reduction of this codec is usually only about 10-30% against simulcast, which makes this technique insufficient for the abovementioned applications. Therefore, in the year 2010, MPEG expert group of ISO/IEC has announced a "Call for Proposals on 3D Video Coding Technology" [2]. In this Call for Proposals (CfP), variants of "multiview video plus depth" are considered as transmission formats. The presence of depth information opens new possibilities to efficiently exploit redundancy between encoded views of the multiview video and, as a consequence, increase the compression efficiency. This paper also continues such an approach.

On the other hand, experts from the ISO/ITU JCT-VC group intensively develop a new video compression standard, the High Efficiency Video Coding (HEVC) [3]. As a result, significant number of new coding tools has been introduced in the HEVC when compared to the MPEG-4 AVC/H.264. Also, the motion compensated prediction mechanism has been considerably developed.

In this article we investigate the possibilities for future multiview video plus depth coding. A proposed Depth-Based Motion Prediction algorithm for inter-view motion prediction was implemented into a new multiview HEVC-based codec.

II. DEPTH-BASED MOTION PREDICTION ALGORITHM

The concept of Depth-Based Motion Prediction (DBMP) results from the observation that motion information (i.e. motion vectors and reference frame indices) for one view can be predicted from depth and motion information available for a reference view. Since motion fields of neighboring views are highly correlated, application of the DBMP algorithm reduces existing data redundancy between these views. And consequently, the overall compression efficiency of the multiview 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 (EIVD) mode [4,5,6] designed for the MVC codec. In this paper, the DBMP algorithm is adopted to multiview video codec based on the HEVC. The main goal of this paper is to explore the possibilities to increase compression efficiency of this new generation multiview codec in case of compression of multiview video enhanced by depth.

For the sake of brevity, detailed description of DBMP will not be presented in this paper. Please refer to [5,6] for more

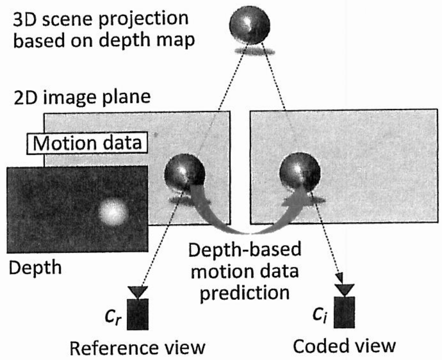

Fig. 1. Point-to-point depth-based motion data prediction.

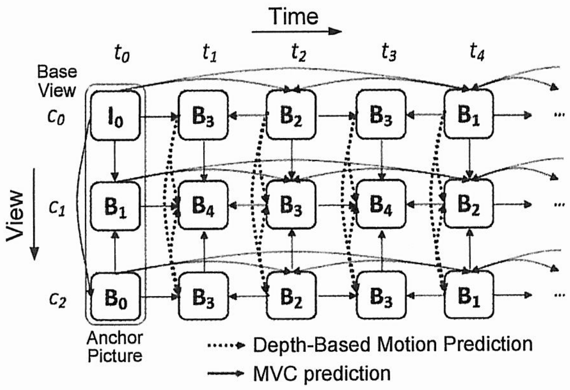

Fig. 2. MVC inter-frame prediction scheme and Depth-Based Motion Prediction (DBMP).

details. In DBMP algorithm, depth information is used to establish a 3D mapping between each pixel in the encoded frame and its counterpart pixel in the reference view (Fig. 1). 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. Consequently, such prediction of motion data for each pixel in encoded view is likely to preserve differences in motion-fields assigned to distinct objects in the scene. As a result, this new depth-based mapping of motion data is more sophisticated and more exact than the 2D mapping used in MVC. Additionally, in DBMP algorithm depth information for encoded view is not required, which makes it applicable for most multiview applications including joint video plus depth coding.

The DBMP algorithm can be utilized in the multiview codec as a motion data prediction method for encoded block. In this block mode, motion information for each pixel in the macroblock is directly inferred from the macroblocks already encoded in the reference view using the DBMP algorithm. As a result, no motion vectors or reference frame indices are transmitted in the bitstream for this macroblock, as the same procedure is repeated in the decoder. This makes the DBMP mode a very efficient tool for encoding motion information, especially for lower bitrates. However, it cannot be applied for the first view in the coding order, which is called the base view. Similarly, the DBMP mode is disabled in case of the anchor pictures as defined in [1], because no motion information referring to other time instances is available in the reference views (Fig. 2).

III. DBMP Prediction for Multiview HEVC

The new idea presented in this paper is to adopt the Depth-Based Motion Prediction algorithm into new generation multiview HEVC-based codec [7]. As described above, the approach used in previous implementations of DBMP-based modes, e.g. the EIVD mode [4,5,6], was based on extending the classic Direct and Skip modes known from MPEG-4 AVC/H.264. However, due to substantial differences in motion compensated prediction between MPEG-4 AVC/H.264 and HEVC, this approach is no longer applicable in HEVC codec, as classic Direct and Skip modes no longer exists in HEVC.

A. *Multiview HEVC*

The multiview HEVC (MV-HEVC) codec is an implementation of compression scheme similar to MVC technology in which the AVC core was substituted by the HEVC coder [7]. As a result, MV-HEVC codec provides additional mechanisms for inter-view prediction known from MVC (Fig. 2) to exploit the inter-view correlation that exists in multiview video and reduce the bitstream representing the side views. As reported by the authors, the coding gains achieved by the MV-HEVC are equal 20-30% when compared to the simulcast HEVC coding.

B. *Implementation of DBMP*

The most efficient modes for encoding a Coding Unit (CU) in HEVC are based on a concept of merging motion information. In this approach, motion vectors and reference frame indices for currently encoded CU are merged with the motion information selected from a set of reference candidates, called merge candidates. Each of the candidates refers to CU placed directly in the neighborhood of the encoded CU. In the current implementation of HEVC [3] all merge candidates are arranged in form of an ordered list (Fig. 3a). The list consists of five candidates: *left*, *top*, *co-located* (the same special position, but different frame), *right-top corner* and *left-bottom corner*. Position on the candidate list is determined based on the probability of the candidate to be chosen for the merging process. This provides a simple but efficient mechanism for manipulating cost of selecting each predictor from the list. If a merge candidate is chosen to represent encoded CU, motion information for this CU is obtained from the candidate. As a result, no motion information or reference frame indices are transmitted in the bitstream for encoded CU.

Considering the abovementioned features of the motion information coding in HEVC, the most elegant and natural way of introducing a new motion predictor into MV-HEVC is extending the existing merge candidate list with proposed DBMP predictor. As a result, original structure of the MV-HEVC codec is preserved, and both number and cost of CU modes remain unchanged. Additionally, DBMP is easily adopted to encoding of wide range of CU sizes without extra syntax modifications.

As presented in Fig. 3b, in proposed implementation, the new DBMP merge candidate is inserted to the candidate list after *left* and *top* candidates, but before *co-located* and two *corner* candidates. This choice of position on the merge candidate list was based on two observations. Firstly, the statistics of merge candidate usage show that first two candidates (*left* and *top*) are selected for CUs covering more

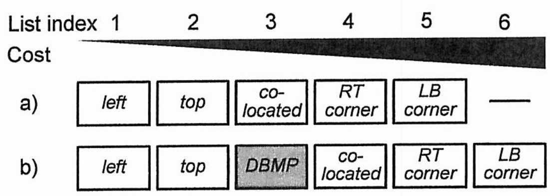

Fig. 3. Merge candidates list: a) original (HEVC), b) modified (new Depth-Based Motion Prediction candidate added).

TABLE I. COMPRESSION PERFORMANCE COMPARISON BETWEEN MV-HEVC+DBMP AND MV-HEVC USING BJONTEGAARD MEASURES

QP 22, 27, 32, 37	2-view case		3-view case			
			Camera c1		*Camera c2*	
Sequence	**ΔPSNRY [dB]**	**ΔBitrate [%]**	**ΔPSNRY [dB]**	**ΔBitrate [%]**	**ΔPSNRY [dB]**	**ΔBitrate [%]**
Poznan Street	0.03	-2.2	0.03	-2.1	0.02	-1.3
Poznan Hall2	0.04	-4.3	0.04	-4.3	0.03	-3.5
Dancer	0.08	-2.8	0.08	-2.9	0.04	-1.4
GT Fly	0.28	-11.7	0.14	-6.2	0.19	-7.4
Kendo	0.17	-6.3	0.15	-5.8	0.10	-3.4
Balloons	0.33	-11.7	0.37	-13.8	0.19	-6.5
Lovebird1	0.03	-1.0	0.02	-0.8	0.01	-0.4
Newspaper	0.15	-5.0	0.15	-5.3	0.06	-1.8
Avg.	0.14	-5.6	0.12	-5.1	0.08	-3.2

than 65% of the picture area. Consequently, they provide a very efficient way of encoding CUs, which should be preserved. On the other hand, our research show that DBMP-based modes in MVC codec are usually selected in 5% to even 80% of macroblocks depending on the method of signaling the modes in the bitstream [8]. Thus, position of the new proposed predictor on the merge candidate list should be possibly high.

IV. EXPERIMENTAL RESULTS

The application of the Depth-Based Motion Prediction algorithm was investigated to assess its impact onto the compression performance of the new generation multiview HEVC-based codec, the MV-HEVC. The evaluation was done by comparing the rate-distortion curves for original MV-HEVC and MV-HEVC with the new DBMP merge candidate implemented (MV-HEVC+DBMP). Both codecs use the HM 3.0 reference software [9] for the HEVC core implementation.

Presented results were obtained using experimental setup described in [2]. Two cases of multiview sequence encoding were investigated: 2-view and 3-view, with both anchor and non-anchor reference pictures for inter-view prediction. Eight standard multiview test sequences were used: "Balloons", "Kendo", "Newspaper", "Lovebird1", "Dancer", "GT Fly", "Poznan Street" and "Poznan Hall2".

Table I presents improvement in compression performance obtained for MV-HEVC+DBMP against MV-HEVC using the modified Bjontegaard measures based on a piecewise cubic interpolation calculated for quantization parameter QP = {22, 27, 32, 37} [10]. These measures reflect average changes of luma PSNR (PSNRY) and bitrate of the encoded dependent view. In 2-view case only one dependant view exists (a non-base view), in 3-view case there are two dependant views, corresponding to cameras *c1* and *c2* in Fig. 2. Bitrate for encoding the depth information is not included as assumed to be transmitted for other purposes with the base layer.

The results show that proposed implementation of DBMP algorithm always improves coding efficiency of the HEVC-based multiview video codec (Table I). The average bitrate savings of MV-HEVC+DBMP against MV-HEVC are about 5.6% for 2-view case. In 3-view case the average bitrate reduction is 5.1% for camera *c1* and 3.2% for camera *c2*. Coding gains are especially noticeable for sequences with complex camera motion (e.g. "GT Fly") or non-translational motion of objects in the scene (e.g. "Balloons"). We can also observe that bitrate savings for camera *c2* in 3-view case are visibly smaller when compared to other test cases. The reason is a smaller correlation between motion fields in coded and reference views which results from the doubled camera distance between these two views.

The more in-depth analysis of both coders reveal the main causes of observed coding gains. Fig. 4 shows the difference between usage of coding units (CUs) encoded with different modes. We can distinguish four types of CU modes in the HEVC-based codec: *Skip*, *Inter-merge*, *Inter* and *Intra*. *Skip* modes are the lowest-cost modes as no motion vectors or texture residuum are encoded for CU. In current implementation of HEVC all CUs in *Skip* mode use merge candidate predictors to represent motion information. *Inter* and *Inter-merge* modes utilize motion compensation to encode CU, however, in the latter case, motion information is represented by merge candidate predictors. *Intra* modes use intra-frame prediction only. As presented in Fig. 4, usage of the low-cost *Skip* modes in MV-HEVC+DBMP multiview codec increases by almost 5% when compared with MV-HEVC in all tested cases. This obviously leads to bitstream reduction.

Another reason for better compression performance of MV-HEVC+DBMP is increasing the amount of CUs with the biggest available size, equal 64×64 pixels (Fig. 5). Consequently, the number of CUs encoded trough division of base 64×64 CU size into smaller parts decreases by almost 3-5%, reducing the bitstream. This reveals an important feature of the DBMP algorithm. When CU is encoded using the DBMP for prediction of motion information, each pixel of CU derives its motion vectors and reference indices regardless of other pixels in this CU. This way, for some CUs, availability of DBMP predictor preserves these CUs from further division into smaller parts as DBMP already provides separate motion data for every pixel in CU.

Fig. 6 shows the average usage of each available merge candidates in MV-HEVC and MV-HEVC+DBMP codecs. Surprisingly, although *DBMP* candidate has bigger cost than *top* candidate, it is used more frequently. This leads to conclusion that position of *DBMP* candidate on the merge candidate list is sub-optimal and should be shifted towards the beginning of the list.

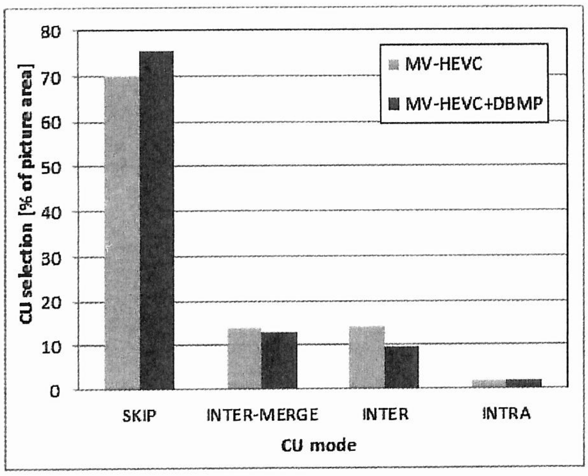

Fig. 4. Usage of coding units (CUs) with different modes in MV-HEVC and MV-HEVC+DBMP codecs.

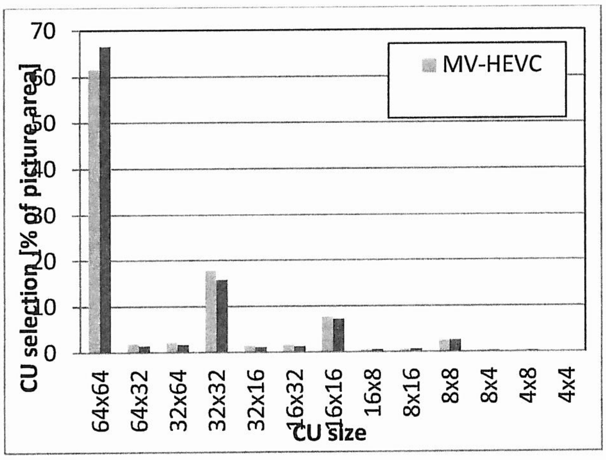

Fig. 5. Usage of coding units (CUs) with different sizes in MV-HEVC and MV-HEVC+DBMP codecs.

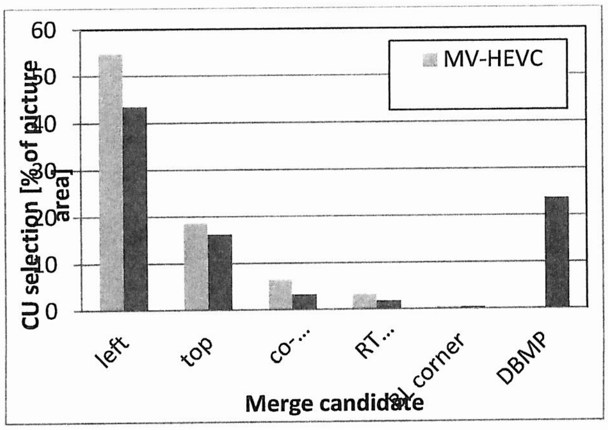

Fig. 6. Usage of coding units (CUs) with different merge candidates in MV-HEVC and MV-HEVC+DBMP codecs.

V. Conclusion

In the paper, a modified technique for the depth-based inter-view prediction of motion vectors is proposed for efficient encoding of multiview video using the new generation codec with HEVC core. This Depth-Based Motion Prediction technique was implemented as a new merge candidate on the merge candidate list of multiview HEVC codec and was designed for depth-enriched multiview video applications, including especially the joint video+depth coding scenarios.

Experiments show that proposed DBMP merge candidate clearly improves compression efficiency of multiview video coding with multiview implementation of HEVC. Results show that adding the proposed DBMP merge candidate to the low-cost modes set of the codec decreases the number of coding units (CU) for which transmitting an additional motion information is required. Additionally, the usage of CUs with smaller sizes decreases as the DBMP allows independent selection of motion information to each point of encoded CU. Consequently, the coding performance increases, resulting in average bitstream reduction of 3.2-5.6% depending on analyzed test case. However, the relative coding gains depend on complexity of motion in encoded content. Also, bitrate savings grow for lower bitrates, as the cost of encoding motion information is higher for small bitstreams.

Moreover, results show that presented implementation of DBMP algorithm in multiview HEVC is sub-optimal. As a consequence, the proposed method has the potential to further increase coding performance of multiview HEVC codec. To conclude, the proposed DBMP merge candidate implements new motion data inter-view prediction technique which may be successfully adopted into developed new generation multiview HEVC-compatible codecs for depth-enriched multiview video.

Acknowledgments

This work was supported by the public funds.

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, MPEG, Doc. N12036, Geneva, Switzerland, March 2011.

[3] "WD3: Working Draft 3 of High-Efficiency Video Coding", ISO/ITU JCT-VC, Doc. JCTVC-E603, Geneva, Switzerland, March 2011.

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

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

[7] M. Domański et al., "Multiview HEVC – experimental results", ISO/ITU JCT-VC, Doc. JCTVC-G582, Geneva, Switzerland, Nov. 2011.

[8] J. Konieczny, M. Domański, "Depth-Enhanced Compression for 3D Video", VCIP 2011, Tainan, Taiwan, Nov. 2011.

[9] https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-3.0rc2/.

[10] F. Bossen, "Common test conditions and software reference configurations", ISO/ITU JCT-VC, Doc. JCTVC-F900, Geneva, Switzerland, July 2011.