DEPTH-BASED INTER-VIEW PREDICTION OF MOTION VECTORS FOR IMPROVED MULTIVIEW VIDEO CODING

Jacek Konieczny, Marek Domański

Chair of Multimedia Telecommunication and Microelectronics
Poznań University of Technology

ABSTRACT

The paper deals with efficient exploitation of mutual correlation that exists in motion fields of individual views in multiview video. The paper describes a new technique for efficient representation of motion data in multiview video bitstreams that carry also depth maps. These depth maps may be used in order to derive motion information from neighboring views. Such inter-view prediction of motion vectors is the core idea of Inter-View Direct compression mode that is proposed in this paper. Application of the new mode yields bitrate reductions between 2% and 13% depending on an individual test sequence, compression scenario and variant of the state-of-the-art multiview compression reference technique. This improvement has been demonstrated by extensive experimental tests that used standard multiview test video sequences.

Index Terms **—** inter-view correlation, motion vector prediction, multiview video coding, depth-enhanced video.

1. INTRODUCTION

Recently, the widely used video compression standard MPEG-4 AVC/H.264 has been extended onto multiview video coding. This extension is known as Multiview Video Coding (MVC) [1]. The key feature of the MVC compression technology is exploitation of mutual correlation between views whenever it reduces the total bitrate needed to deliver video for all the views. Unfortunately, MVC does not satisfy all the needs related to 3D video compression. For future 3D video coding technology, there exists an opinion that advantageous would be some backward compatibility with existing MPEG-4 AVC / H.264 video compression standard as well as with its multiview (MVC) extension. Therefore, research on extensions of multiview video coding is very important for future applications.

The future 3D video transmission technology would likely comprise also the tools for delivery of depth information together with camera parameters. When this additional information is already available in an encoded video bitstream, it should be exploited for compression as well. Therefore, compression of such depth-enriched video should better remove inter-view information redundancy.

This paper describes a new technique for efficient representation of motion data in depth-enriched multiview video bitstreams. Such compression benefits from the observation that motion information is often similar for different views of the same scene. The idea is to improve compression of some views by reusing appropriate motion vectors from the reference views. In the course of MVC standardization, similar concepts were already investigated [2]. In particular, for MVC, proposed was an additional compression mode called Motion Skip (MS) [3]. In this mode, for a currently encoded macroblock, a global disparity vector is used for initial identification of the corresponding motion vectors from the reference frames. Actually, the motion vectors that are found in the optimization that starts at the position given by the global disparity vector are reused from the reference frames. Application of Motion Skip mode increases compression performance of the respective views [2] at the cost of additional complexity of the encoder. Later, versions of Motion Skip mode that use better accuracy of global disparity than originally described were also introduced [4,5].

In contrary to the abovementioned proposals, here, we are going to propose a technique that exploits local disparity in order to identify the motion vectors that may be reused from the reference frames. We are going to show that such an approach leads to even better compression performance and is not computationally complex.

2. MAIN IDEA

Assume that multiview video is transmitted together with respective depth information. It is reasonable to assume that for individual views depth information is represented by disparity maps as well as camera parameters.

The main idea is to calculate motion information for some views from the motion information available for other views called reference views. Therefore motion information does not need to be transmitted for some views. Obviously, such reduction of transmitted motion information results in improved compression efficiency, especially for lower bitrates.

The abovementioned means of motion information representation should be considered as an additional mode of video compression. Its prospective selection would be decided in the process of rate-distortion optimization in an encoder.

This new compression mode will be called Inter-View Direct (IVD) mode on the analogy of classic Direct mode. In both these modes, motion vectors are not transmitted in a bitstream but rather obtained from the reference picture. In classic Direct mode, the reference pictures are in past and maybe in future. In the newly proposed Inter-View Direct mode, the reference pictures are in other views.

For Inter-View Direct mode, disparity define a mapping such that each pixel in the encoded frame has its counterpart pixel in the reference frame. In the reference frame, except of intra-coded macroblocks, a pixel belongs to a macroblock or its partition with an uniquely defined motion vector and the corresponding temporal reference frame. Using the mapping yielded by the disparity map, a motion vector can be obtained for each pixel in the encoded image.

3. THE ALGORITHM

The proposed Inter-View Direct (IVD) mode is a new video encoding mode that may be selected for a macroblock. As

reference frames one or more frames from the neighboring views may be chosen.

Using view c_0 (see Fig. 1) as the reference view, the IVD mode may be used for both c_1 and c_2 views. Otherwise, having frames from views c_0 and c_2 already encoded, the respective frame from c_1 may be encoded with IVD mode having reference frames from views c_0 and c_2.

If Inter-View Direct mode is selected for a macroblock, its motion information, such as motion vectors and reference indices, is directly inferred from the already encoded macroblocks in the neighboring views. For such macroblock, no motion information is transmitted in the bitstream. In the decoder, the motion information is inferred from the neighboring view in the same way as in the encoder. However, IVD mode should be disabled in case the currently encoded macroblock is in a picture of base view or an anchor picture as defined in JMVM [6]. In such cases, IVD mode cannot be applied because no reference view or motion information are available (Fig. 1).

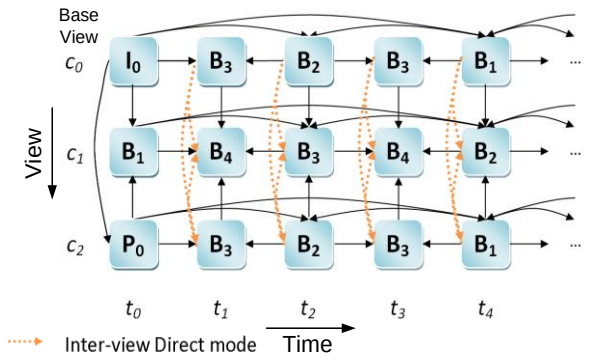

Fig. 1. MVC+IVD mode prediction schemes.

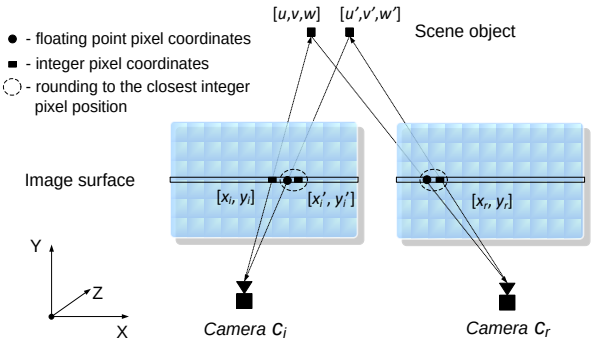

Fig. 2. Multiview point correspondence.

In the description of the algorithm we denote the current view as c_i, i.e. c_i denotes the view related to the currently encoded picture. Similarly, the reference view is c_r.

For the case of accessibility of depth maps (or disparity maps) for both coded and reference views, the algorithm is described as follows (Fig. 2):

1) Project the pixel location (x_i,y_i) in the coded picture from view c_i into world coordinates (u,v,w) using transformation:
$[u,v,w]^{\mathrm{T}} = \mathbf{R}(c_i)\mathbf{A}^{-1}(c_i)[x_i,y_i,1]^{\mathrm{T}}\mathbf{D}(x_i,y_i,t,c_i) + \mathbf{T}(c_i)$,
where $\mathbf{A}(c)$ is an intrinsic camera parameter matrix of camera c, $\mathbf{R}(c)$ is a rotation matrix of camera c, $\mathbf{T}(c)$ is a translation vector of camera c and $\mathbf{D}(x,y,t,c)$ is the depth value related to camera c view at pixel (x,y) and time instance t.
2) Map the world coordinates (u,v,w) into local coordinates of the reference view picture c_r using:
$[x_r \cdot z_r, y_r \cdot z_r, z_r]^{\mathrm{T}} = \mathbf{A}(c_r)\mathbf{R}^{-1}(c_r)\{[u,v,w]^{\mathrm{T}} - \mathbf{T}(c_r)\}$.
The corresponding pixel coordinates of the reference view picture are (x_r,y_r) and z_r is a depth value at point (x_r,y_r).
3) Test the corresponding pixel positions (x_i,y_i) and (x_r,y_r) for occlusion. If the coded pixel (x_i,y_i) is not occluded and the corresponding pixel (x_r,y_r) has all required information for inter-frame prediction mode available, the pixel position (x_r,y_r) can be assigned as a source of motion information for the coded pixel (x_i,y_i). Otherwise, derivation of motion information from the selected image point (x_r,y_r) in the reference view is not possible and point (x_i,y_i) from the encoded macroblock remains unassigned. In such a case another reference view should be inspected according to the inversed view coding order.
4) If no other reference view is available, motion information for pixel (x_i,y_i) is derived from the neighboring points of the same view, for which the corresponding pixel in the reference view was successfully found. For example, an inpainting algorithm might be used for that purpose. Also the closest neighbor from background or foreground may be chosen. However, this issue is still a subject of further investigation.
5) If the method described in *step 4* cannot be applied, another possibility to gain the motion information for pixel (x_i,y_i) is to use a standard Direct mode prediction, i.e. to use intra-view prediction of motion vectors instead of the abovementioned inter-view prediction of motion vectors.
6) When the source of motion information for each point (x_i,y_i) of the coded macroblock is finally determined, motion compensation is done independently for each point according to the derived motion vectors and reference indices.

Application of the proposed IVD mode increases the encoder complexity only marginally. The only information which is derived from the reference view is the data originally encoded in the bitstream and there are no additional calculations or restrictions to encoding process for the reference view. The only computational overhead in the encoder/decoder is related to the depth-based rendering and occluded points procedures, thus resulting in only very slight increase of encoding time.

4. IMPLEMENTATION OF THE ALGORITHM

In order to evaluate Inter-view Direct mode, it was integrated into MPEG-4 MVC reference software model [1] and appropriate minor modifications have been made in the multiview bitstream syntax and semantics.

1. In order to notify the use of IVD mode to the decoder, a new flag *ivd_flag* is included in the bitstream. This new mode is signaled with modification to the existing Direct mode macroblock layer syntax in non-anchor pictures of non-base view. With the proposed change, an additional bit is added if *mb_type* is signaling the Direct mode selection. The additional flag *ivd_flag* allows to distinguish the new IVD mode from the traditional Direct mode. Flag *ivd_flag* = 1 signals IVD mode, otherwise the conventional Direct mode is used.
2. The IVD mode is disabled if the currently encoded macroblock is in the picture of the base view or is an anchor picture. Encoding of the base view is fully compliant with MPEG-4 AVC / H.264.
3. For the experiments, no dedicated context model for the *ivd_flag* was implemented in the CABAC encoder, and consequently the *ivd_flag* encoding is sub-optimal.
4. In order to reduce complexity, the filling algorithm for occluded points is based on arbitrary selection of the closest neighbors. However, application of more sophisticated voting algorithm should improve the overall compression efficiency.
5. In the described version of IVD mode, disparity maps are required for both the reference and encoded views. In that way, only minor modifications are needed in the reference MVC software model. Nevertheless, the IVD mode is usable with a

disparity map for the reference view only, and this issue will be investigated in another work of the authors.

6. View coding order is determined from the coding order included in the MVC bitstream.
7. Depth-based rendering requires intrinsic and extrinsic camera parameters. In MVC, this information is provided in Multiview Acquisition Info SEI messages [7,8]. However, if the depth information is available as disparity maps, the *z_near* and *z_far* values are needed in order to convert disparity to depth. These two parameters need to be encoded into the bitstream for each view as it was implemented in the authors' extension of Multiview Scene Info SEI.
8. The current experimental implementation is using IVD mode for 16x16-pel macroblock partitioning only. Obviously, IVD mode is applicable also to smaller blocks.

5. EXPERIMENTAL RESULTS

The goal of the experiments was to assess the impact of application of the new IVD mode onto the compression performance for multiview video . It was done by comparing the rate-distortion lines for standard MVC codec to the MVC augmented with Motion Skip mode and Inter-View Direct mode.

The Inter-View Direct mode was implemented into the JMVC 4.0 [9] software (JMVC+IVD). Comparisons have been made with respect to original JMVC 4.0 software (JMVC) as well with respect to JMVM 8.0 [10] software with enabled motion skip and single loop decoding (JMVM+MS). For each multiview video, all views are coded with hierarchical B frames as shown in Fig. 1.

As the analyzed implementation of IVD mode requires depth map images both for the reference view and the encoded view and most of the currently available reference depth map collection contains depths for only two views, two scenarios of stereo sequence encoding are included in the presented results. *Scenario 1* refers to the case of encoding view c_2 in Fig. 1 with the c_0 view as the reference view – only anchor reference pictures of one reference view are available. In *Scenario 2* view c_1 in Fig. 1 is encoded, however, with only c_0 available as the reference view. In this case both anchor and non-anchor reference pictures are used.

The experiments used 5 multiview test sequences: "Book Arrival" [11], "Newspaper" [12], "Lovebird1" [13], "Champagne tower" [14] and "Pantomime" [14]. The sequences were encoded according to MVC hierarchical coding structure with GOP size of 12. The experiments were done for quantization parameter QP = 24, 30, 36 or 42.

For decoded video, its quality was measured by luma Peak Signal-to-Noise Ratio (PSNRY) that was averaged over first 96 frames from each test sequence. The proposed technique is aimed at applications where depth information is send anyway for some other purposes. Therefore the bitrates are calculated for views only, i.e. without bits needed for transmission of depth information. In fact, the bitrate values are calculated for one view only.

The Bjontegaard measures [15] have been used in order to express improvement in compression performance obtained for the codec with Inter-View Direct mode (JMVC+IVD) as compared to the standard MVC codec without Motion Skip mode (JMVC) and with Motion Skip mode (JMVC+MS). Positive values of *ΔPSNR* denote average increase of PSNRY for JMVC+IVD as compared to other codecs. Similarly, negative values of *ΔBitrate* denote average decrease of bitrate for JMVC+IVD as compared to other codecs. Rate-distortion lines for selected sequences are shown in Fig. 3 and Fig. 4, for both scenarios respectively.

The proposed IVD mode always improves compression efficiency of multiview video codecs. The average bitrate saving is about 6.5 % for *Scenario 1* and about 4.9% for *Scenario 2* (see comparison JMVC+IVD against JMVC in Table 1 and 2, respectively). The results prove that, in average, Inter-View Direct mode improves multiview compression performance better than Motion-Skip mode. When compared with

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

QP 24,30,36,42	*JMVC*		*JMVM+MS*	
	ΔPSNR [dB]	*ΔBitrate [%]*	*ΔPSNR [dB]*	*ΔBitrate [%]*
Book Arrival	0.35	-9.6	0.49	-13.3
Newspaper	0.15	-3.6	0.47	-10.8
Lovebird1	0.15	-4.7	0.47	-13.7
Champagne tower	0.29	-6.5	0.37	-8.3
Pantomime	0.38	-8.0	0.40	-8.3
Average	0.27	-6.5	0.44	-10.9

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

QP 24,30,36,42	*JMVC*		*JMVM+MS*	
	ΔPSNR [dB]	*ΔBitrate [%]*	*ΔPSNR [dB]*	*ΔBitrate [%]*
Book Arrival	0.23	-6.5	0.11	-3.2
Newspaper	0.11	-2.7	0.21	-5.0
Lovebird1	0.09	-2.9	0.34	-10.3
Champagne tower	0.24	-5.3	-0.01	0.2
Pantomime	0.33	-7.3	-0.25	5.8
Average	0.20	-5.0	0.08	-2.5

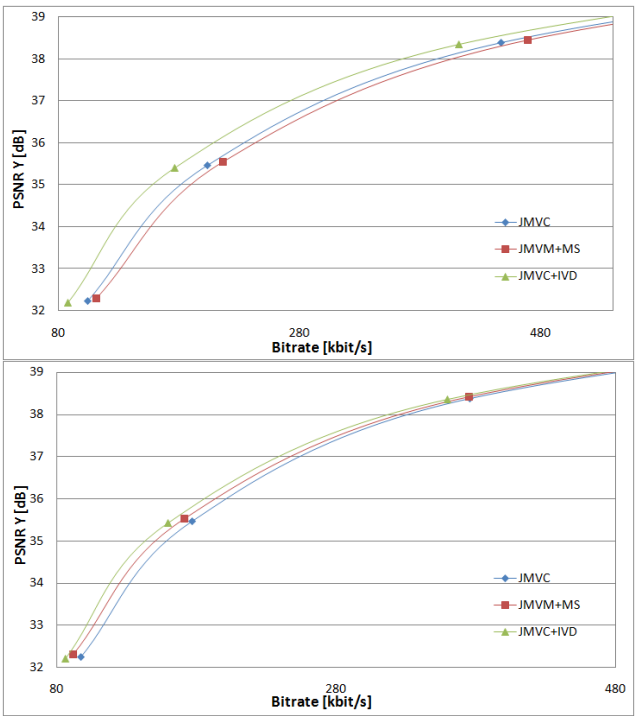

Fig. 3. Results of JMVC, JMVM+MS and JMVC+IVD coding for "Book Arrival": top – *Scenario 1*, bottom – *Scenario 2*.

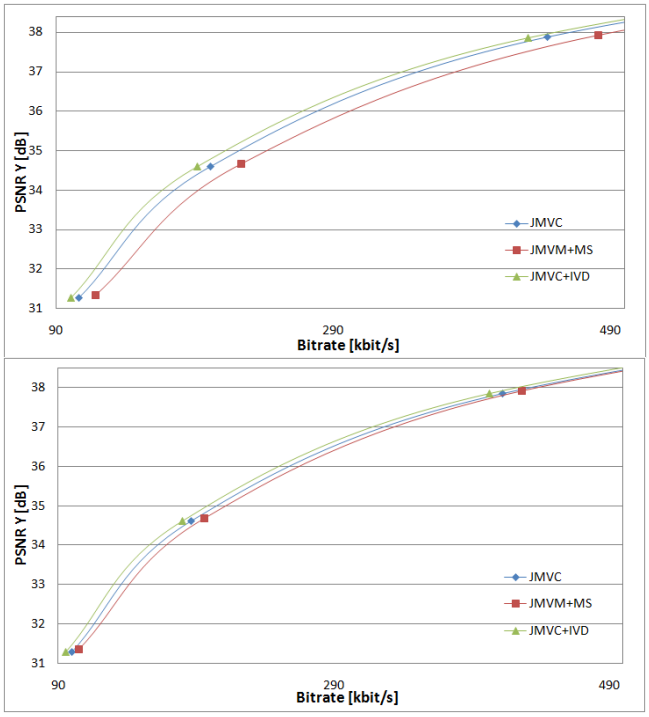

Fig. 4. Results of JMVC, JMVM+MS and JMVC+IVD coding for "Newspaper": top – *Scenario 1*, bottom – *Scenario 2*.

JMVM+MS, the average bitrate savings of JMVC+IVD are about 10.8% for *Scenario 1* and 2.5 % for *Scenario 2*. However, for some sequences in *Scenario 2*, the comparison with JMVM+MS shows that the proposed Inter-View Direct mode improves compression performance less than Motion Skip mode. This is noticeable especially for the "Pantomime".

The overall compression efficiency for *Scenario 2* is better than for *Scenario 1* for all tested cases due to unused inter-view correlation in original MVC scheme for *Scenario 1* (Fig.1). For all test sequences in *Scenario 1*, the compression efficiency of JMVM+MS is worse than for the JMVC. In contrast, JMVC+IVD performs much better than JMVM+MS for *Scenario 1*, always giving significant bit-rate savings.

Apart from the observations presented above, it should be emphasized that JMVC+IVD compression performance depends on the quality of the depth maps. Probably, poor quality of depth maps for *Pantomime* sequence is a reason of reduced efficiency of IVD mode for this sequence.

6. CONCLUSION

In the paper, new technique for inter-view prediction of motion vectors is proposed for efficient encoding of multiview video. This new technique is proposed as a core tool for a new Inter-View Direct mode of macroblock compression for applications in depth-enriched multiview video compression.

Careful experiments show that this newly proposed mode clearly improves compression efficiency of multiview video coding. Embedding of this mode into the state-of-the art multiview video codec usually results in an increase of bitrate of less than 10%. Probably, availability of more accurate depth maps could increase the importance of this new mode. The proposed improvement causes only negligible increase of encoder and decoder complexity.

Moreover, the newly proposed Inter-View Direct mode usually improves compression performance more than Motion Skip mode known from the references. This difference would probably grew as more accurate depth maps become available.

Compression performance of multiview video coding with Inter-View Direct mode may be further improved by application of this new mode to small blocks as well.

The above mentioned issues imply that proposed technique exhibits some potential in compression of depth-enriched multiview video that will likely be an important issue for 3D video technology.

ACKNOWLEDGMENTS

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

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] A. Vetro, P. Pandit, H. Kimata, A. Smolic, "Joint Multiview Video Model (JMVM) 5.0", ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-X207, July 2007.

[4] 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.

[5] "Joint multiview video model (JMVM) 8.0", JVT-AA207, Geneva, Switzerland, April 2008.

[6] 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.

[7] 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.

[8] 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.

[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.