INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG/M17800
July 2010, Geneva, Switzerland

Source	**Poznań University of Technology**
Status	**Technical Contribution to 3D Video Coding**
Title	**Inter-View Direct Mode for Multiview Video Coding**
Author	**Jacek Konieczny** (jkonieczny@multimedia.edu.pl) and **Marek Domański** (domanski@et.put.poznan.pl) Poznań University of Technology, Chair of Multimedia Telecommunications and Microelectronics, Poznań, Poland

1 Introduction

This document presents a new Inter-View Direct mode (IVD) for multiview video coding which originates from the idea that motion fields of neighboring views in multiview sequence are highly correlated. Proposed technique provides an efficient representation of motion data in multiview video bitstreams that carry also depth/disparity maps. In the proposed method, the motion information, such as motion vectors and reference indices, for each pixel of encoded mackroblock is directly inferred from the already encoded macroblocks in the neighboring views at the same temporal instance. Thus, the proposed method is analogy of classic Direct mode as motion vectors and reference indices are not transmitted in a bitstream but obtained from the reference view. Since the disparity between encoded and reference view exists, the IVD mode uses additional information about mapping of each pixel from encoded view into reference view represented in form of disparity maps. Preliminary experimental results show the bitrate reduction up to 10% in comparison with MVC.

2 Proposed method

2.1. Inter-View Direct mode

We propose a novel approach to efficient representation of motion data in multiview video bitstreams that carry also depth/disparity maps. In this approach motion information for encoded view is directly inferred from depth and motion information available for reference views. Consequently, motion information encoded in the bitstream can be reduced. This proposal is based on the approach introduced in [1] and is also described in another authors paper [2].

In the proposed Inter-View Direct mode (IVD), depth information is used to define a mapping between each pixel in the encoded frame and its counterpart pixel in the reference frame. With this mapping motion vectors and reference indices 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. We assume that depth information is available at the decoder, regardless of the way of delivering it. In proposed approach depth information associated only with reference views is used, which makes the method suitable also for joint depth and video coding.

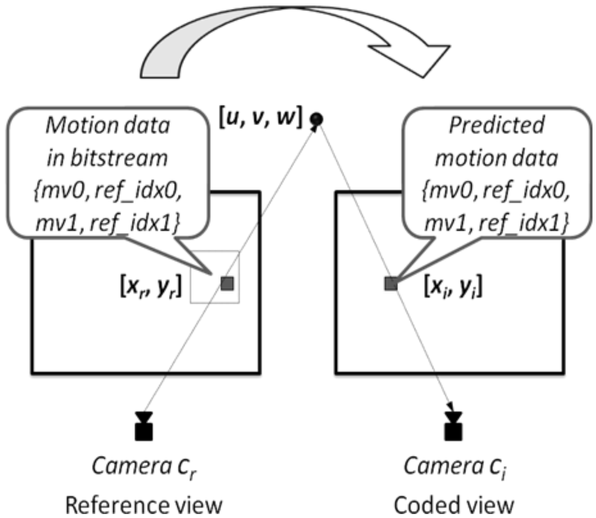

Fig. 1. Derivation of motion information from corresponding point in reference view.

The algorithm of determining the source of motion information for each point (x_i,y_i) of encoded mackroblock in view c_i is as follows (Fig.1):

1. Set reference view c_r := first available reference view for c_i according to inversed view coding order.
2. Use Depth-Image Based Rendering technique to project each pixel location (x_r,y_r) in reference view c_r into coded view c_i to determine corresponding pixel location (x_i,y_i) in view c_i.
3. Test corresponding pixel positions (x_i,y_i) and (x_r,y_r) for occlusion: if the pixel (x_i,y_i) has no corresponding pixel (x_r,y_r) assigned to it after reprojecting all pixel locations from reference view c_r to view c_i – it is occluded.
4. IF (x_i,y_i) is not occluded AND (x_r,y_r) has a motion information assigned: MD(x_i,y_i) := MD(x_r,y_r), where MD(x,y) is motion information assigned to point (x,y).
   ELSE c_r := next reference view for c_i according to inversed view coding order; GOTO *Step* 2.
5. IF (x_i,y_i) has no motion information assigned MD(x_i,y_i) := MD(x_i',y_i'), where MD(x_i',y_i') is motion information of closest neighboring pixel of (x_i,y_i) for which motion information was successfully found.
6. IF (x_i,y_i) has no motion information assigned AND *Step* 5 could not be applied MD(x_i,y_i) := standard Direct mode prediction.

The proposed method is available in the non-base view and non-anchor pictures only. As a result the base view remains fully compliant with MPEG-4 AVC / H.264. The order of reference view inspected during the search of corresponding pixels of encoded macroblock is equal to inversed view coding order included in the MVC bitstream with respect to all views encoded before the currently encoded view. Thus no additional syntax is required for referencing the inter-view pictures. The algorithm uses only reference view depth/disparity information which makes it suitable for all depth-enriched applications including joint depth and video coding.

2.2. Syntax

The use of IVD mode is signaled to the decoder using a new flag *ivd_flag* included in the bitstream. This new flag is added into bitstream as a 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 after the *mb_type* if *mb_type* is signaling the Direct mode selection in non-anchor pictures of non-base view. 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.

Depth-Image Based Rendering requires intrinsic and extrinsic camera parameters. In MVC, this information is provided in Multiview Acquisition Info SEI messages [3,4]. However, if the depth information is available in form of disparity maps, the *z_near* and *z_far* values are needed in order to convert disparity to depth. In our proposal we encoded these two parameters into bitstream as an extension of Multiview Scene Info SEI.

3 Experimental results

The IVD mode was implemented into the JMVC 4.0 software (JMVC+IVD) [5] and compared with the codecs implemented using:

1) original JMVC 4.0 software (JMVC) [6],
2) JMVM 8.0 software with enabled motion skip and single loop filtering (JMVM+MS) [7].

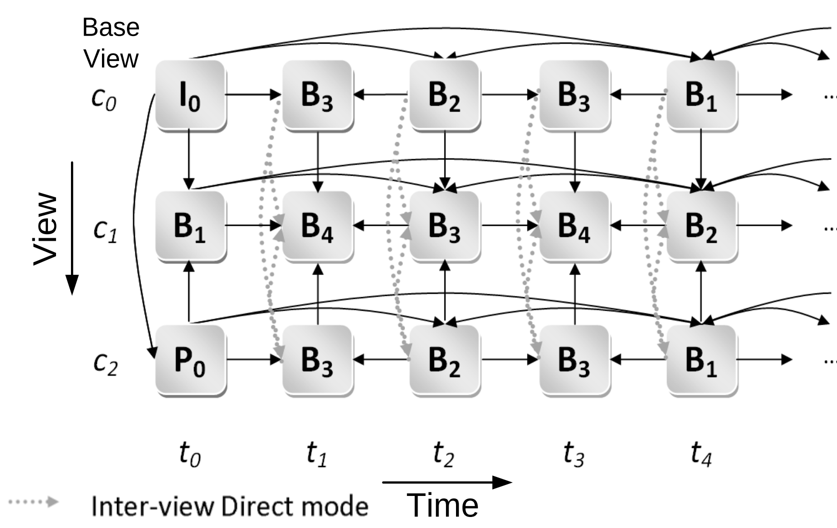

Fig. 2. MVC+IVD mode prediction scheme.

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

The results were obtained using 5 standard multiview test sequences: “Book Arrival” [8], “Newspaper” [9], “Lovebird1” [10], “Champagne tower” [11] and “Pantomime” [11]. 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.

The results are presented in Figs. 3-7. Table 1 and 2 present improvement in compression performance obtained for codec with Inter-View Direct mode (JMVC+IVD) against the standard MVC codec (JMVC) and codec with Motion Skip mode (JMVC+MS) using the Bjontegaard measures [12].

Presented bitrates are calculated for encoded view only. The proposed technique is designed for applications that use depth information regardless of the method of delivering it. Consequently. bits needed for transmission of depth information are not included (we assume that depth information is transmitted already for other purposes).

As shown in the results, proposed technique always improves compression efficiency of multiview video codec (JMVC). The average bitrate saving is about 6.9% for *Scenario1* and about 5.3% for *Scenario2*. Also, in average, IVD mode improves multiview compression performance better than JMVM+MS. In this case the average bitrate savings of JMVC+IVD are about 11.3% for *Scenario1* and 2.8% for *Scenario2*.

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

QP 24,30,36,42	*JMVC*		*JMVM+MS*	
	ΔPSNR [dB]	*ΔBitrate [%]*	*ΔPSNR [dB]*	*ΔBitrate [%]*
Book Arrival	0.33	-9.2	0.47	-12.9
Newspaper	0.22	-5.2	0.54	-12.3
Lovebird1	0.15	-4.6	0.47	-13.6
Champagne tower	0.32	-7.2	0.40	-9.0
Pantomime	0.40	-8.3	0.42	-8.6
Average	0.28	-6.9	0.46	-11.3

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	0.22	-6.0	0.10	-2.7
Newspaper	0.17	-4.0	0.26	-6.1
Lovebird1	0.09	-2.9	0.34	-10.3
Champagne tower	0.26	-5.9	0.01	-0.4
Pantomime	0.35	-7.7	-0.23	5.4
Average	0.22	-5.3	0.10	-2.8

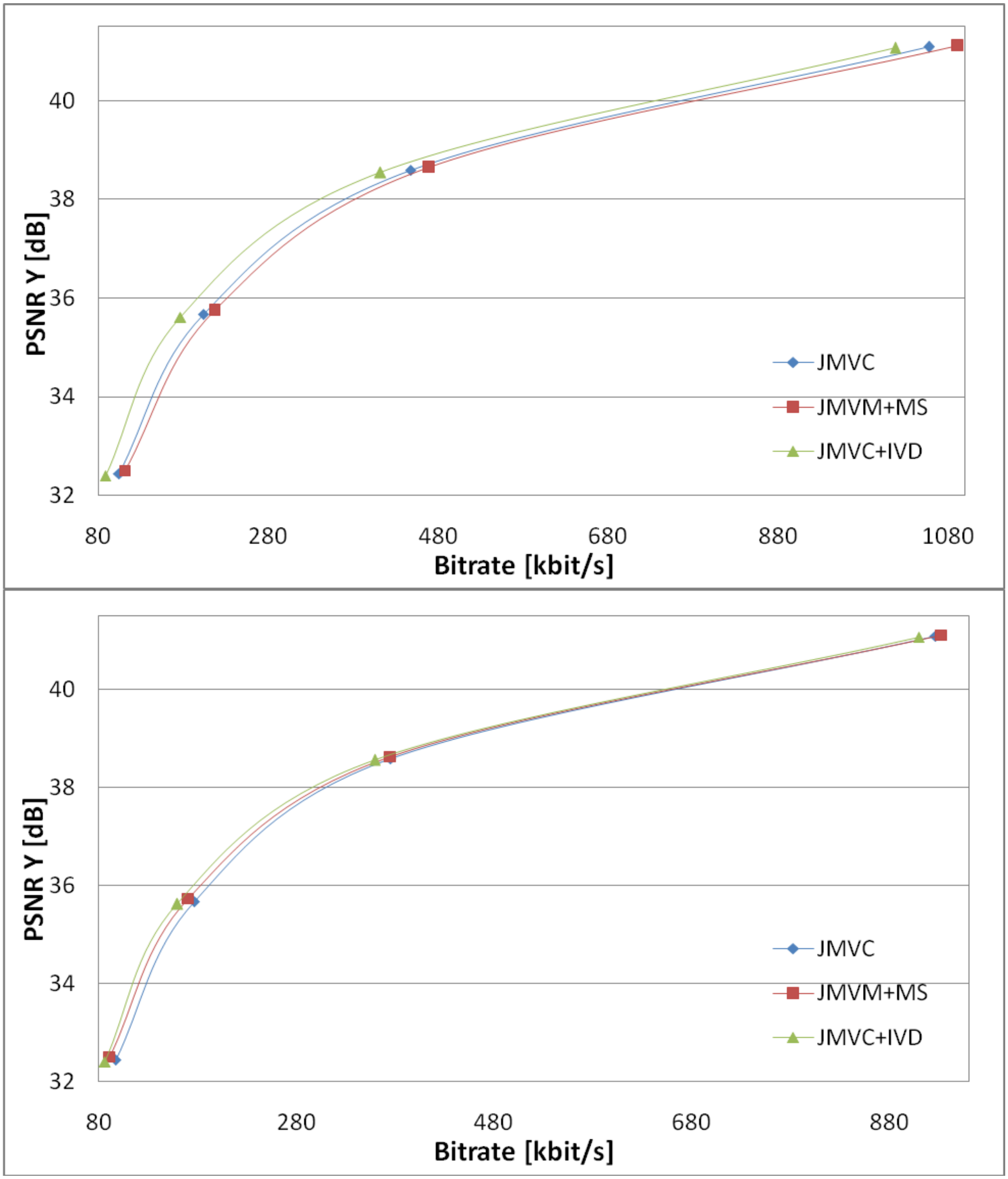

Fig. 3. Rate-distortion curves: top – *Scenario1*, bottom – *Scenario2*, sequence: “Book Arrival”.

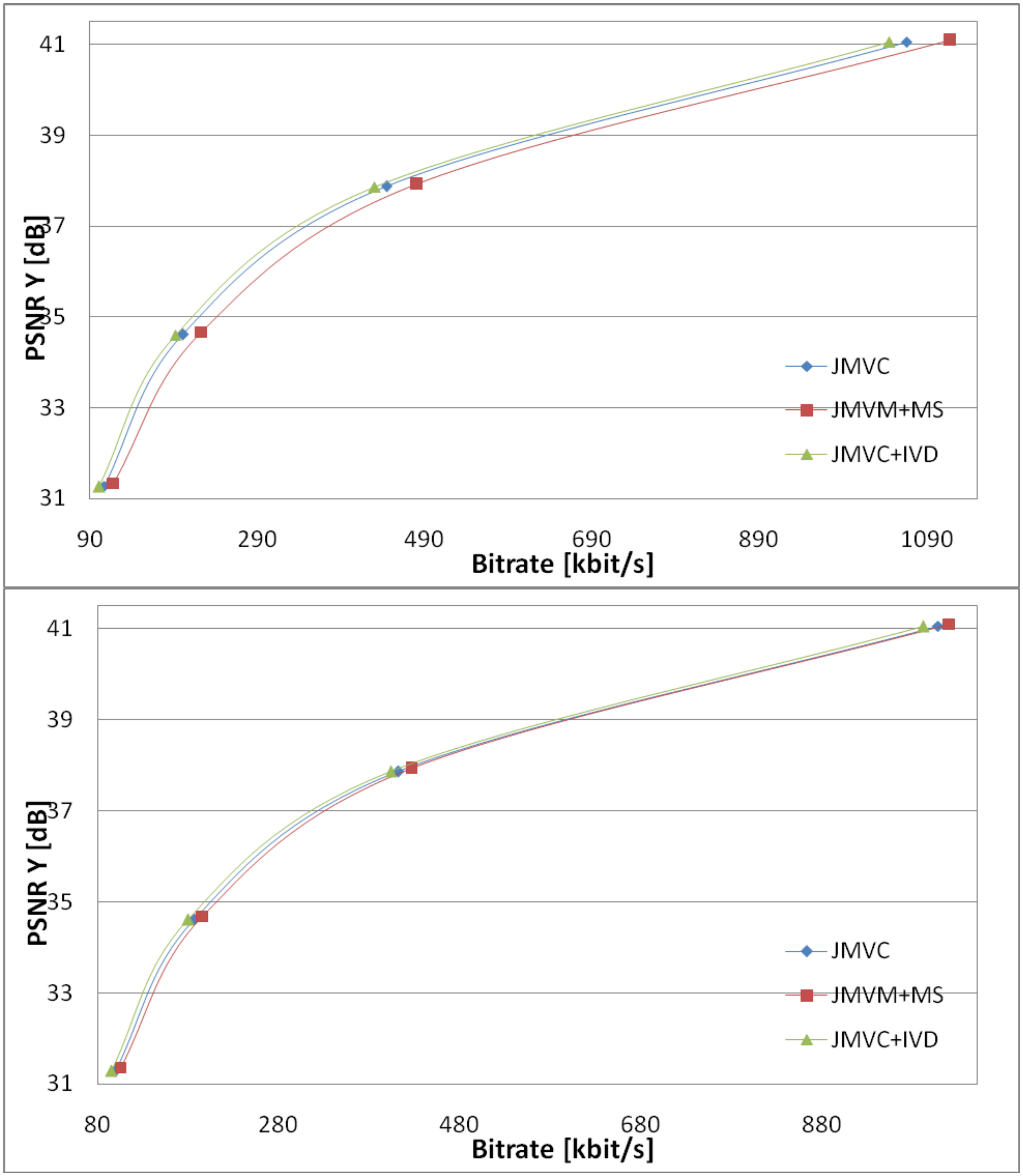

Fig. 4. Rate-distortion curves: top – *Scenario1*, bottom – *Scenario2*, sequence: “Newspaper”.

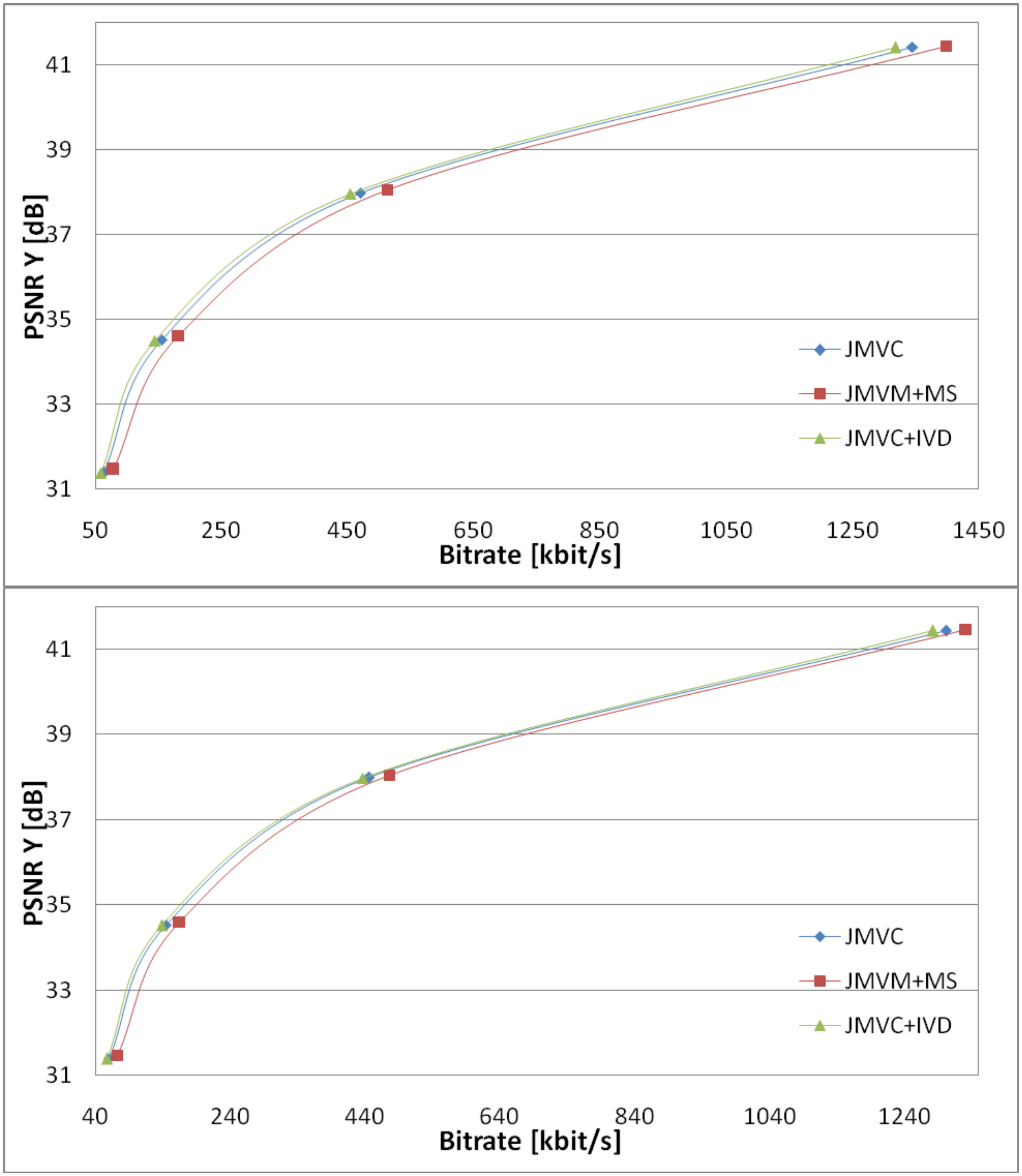

Fig. 5. Rate-distortion curves: top – *Scenario1*, bottom – *Scenario2*, sequence: “Lovebird1”.

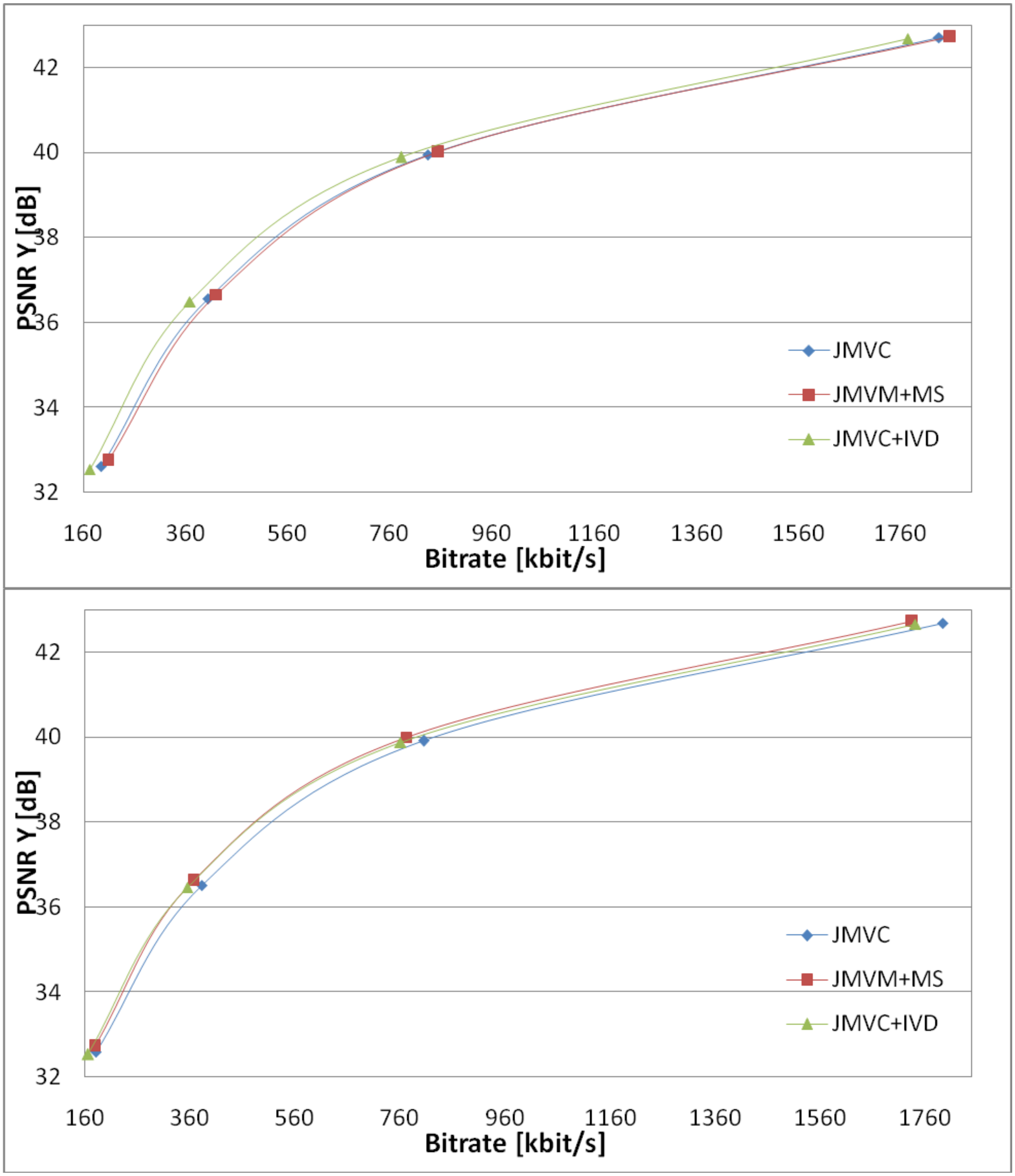

Fig. 6. Rate-distortion curves: top – *Scenario1*, bottom – *Scenario2*, sequence: “Champagne tower”.

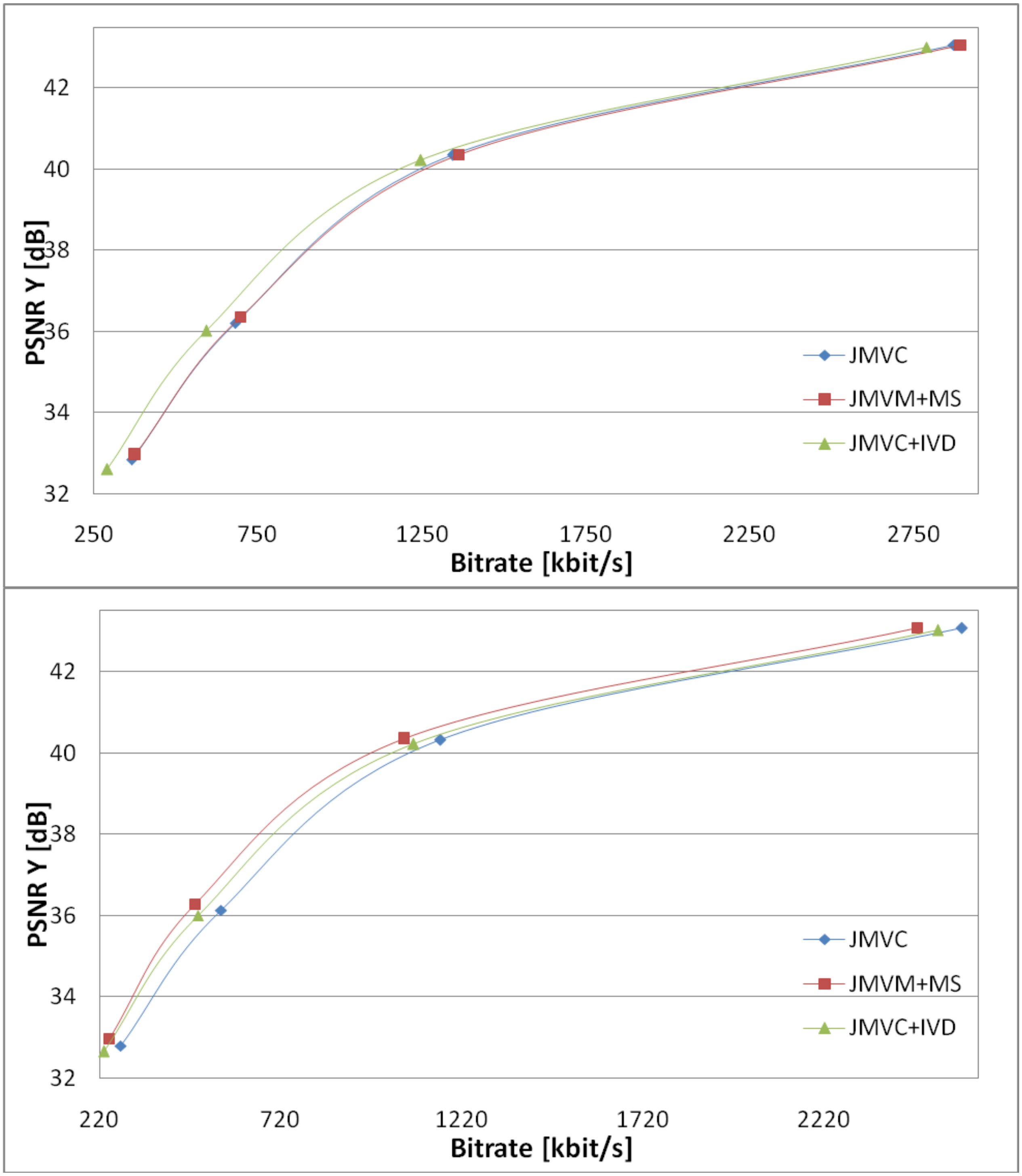

Fig. 7. Rate-distortion curves: top – *Scenario1*, bottom – *Scenario2*, sequence: “Pantomime”.

4 Conclusions

In this document, the Inter-View Direct mode was proposed to improve the performance of multiview video coding. The proposed method eliminates the motion information such as motion vector and reference indices from the bitstream and infers it from already encoded neighboring view. First, for each pixel of encoded mackroblock the corresponding pixel position is located in the neighboring view using the Depth-Image Based Rendering technique and depth information of the neighboring view. Next, motion information is copied from these corresponding pixels in the neighboring view. Experimental results show that the proposed method can significantly improve the coding efficiency for standard test sequences. We recommend adopting this technique into future 3D video coding standard.

5 References

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

[2] J. Konieczny, M. Domański, „Extended Inter-View Direct Mode for Multiview Video Coding", *to be published*.

[3] 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.

[4] 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.

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

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

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

[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 Contribution VCEG-M33, Austin, USA, April 2001.