3D-HEVC EXTENSION FOR CIRCULAR CAMERA ARRANGEMENTS

Jakub Stankowski, Łukasz Kowalski, Jarosław Samelak,
Marek Domański, Tomasz Grajek, Krzysztof Wegner

Poznań University of Technology, Chair of Multimedia Telecommunications and Microelectronics

ABSTRACT

The paper presents an extension of 3D-HEVC for the circular camera arrangements. It generalizes the derivation of the disparity vectors from depth data for sequences captured using cameras located on an arc. The general equation for disparity calculations has been implemented instead of the simplified equation used in 3D-HEVC. Experiments have been performed on widely recognized multiview test sequences with both circular and linear camera arrangements. In the case of sequences obtained using cameras located on an arc, the proposed extension provides about 6% bitrate reduction when compared to 3D HEVC, and does not influence the 3D-HEVC performance for linear view arranged sequences.

Index Terms — 3D-HEVC, FTV, multiview compression, non-linear camera arrangements, circular, arc camera setup

1. INTRODUCTION

Currently 3D-HEVC [1] is the state-of-the-art compression technology for 3D video in "multiview video and depth" format (MVD) [2]. This technology has been developed by Collaborative Team on 3D Video Coding Extensions Development (JCT-3V) formed between ISO/IEC and ITU-T and is foreseen to be incorporated into the HEVC standard (High Efficiency Video Coding) as Annex I of ISO/IEC MPEG-H Part 2 and ITU-T Recommendation H.265.

The 3D-HEVC is built on the top of the MV-HEVC [1] codec. The MV-HEVC utilizes inter-view prediction between the views and no information about scene structure (depth) is used. The MV-HEVC simply uses side views as another source for inter-prediction (in the same way as previous fames are used for motion compensated inter-prediction).

The goal of the 3D-HEVC was to exploit the information about 3D scene structure (in form of depth maps) to increase coding efficiency of 3D video. During the development of the 3D-HEVC technology, explicit 1D parallel views arrangement has been assumed. Therefore, the 3D-HEVC coding efficiency for non-linear camera arrangements is far from optimal.

Many modern Super Multiview (SMV) displays require circular (arc) view arrangements for better user experience. Therefore, efficient compression technology for non-linear (e.g. arc) camera arrangements is of great interest.

In this paper we present extension of 3D-HEVC towards arbitrary view setup, in particular the arc view arrangement.

2. GENERALIZATION OF 3D-HEVC TOOLS

Inter-view prediction which is the core tool of MV-HEVC, in general, does not assume any particular view arrangement. But, in order to increase the coding efficiency of 3D-HEVC, a number of specific tools was introduced and others were simplified with explicit 1D parallel view arrangement assumption.

The most significant simplification states that the views are vertically aligned. Therefore, disparity vectors are restricted only to horizontal direction, and they can be derived from depth data through simple linear equation.

$$\begin{aligned} d_h &= a \cdot v + b \\ d_v &= 0 \end{aligned} \tag{1}$$

where the horizontal component d_h of the disparity vector is calculated from depth sample value v and scaled to disparity by using scale factor a and offset b transmitted for each view, and vertical component of disparity vector d_v is always set equal to zero.

In the proposed 3D-HEVC extension, we have removed restriction only for horizontal direction, and applied general derivation of disparity vector based on depth data.

Based on the point position x, y in the picture and associated depth sample value v, position of the corresponding point x_r, y_r in the reference view can be calculated. Then, derived disparity vector is simply the difference in the position of considered point and the corresponding point in the reference view

$$\begin{aligned} d_h &= x_r - x \\ d_v &= y_r - y \end{aligned} \tag{2}$$

The position of the corresponding point in the reference view can be derived through depth based projection of the point according to the equation (3)

$$\begin{bmatrix} z_r \cdot x_r \\ z_r \cdot y_r \\ z_r \\ 1 \end{bmatrix} = P_r \cdot P^{-1} \cdot \begin{bmatrix} z \cdot x \\ z \cdot y \\ z \\ 1 \end{bmatrix}, \tag{3}$$

where P_r, and P are projection matrices for both the reference view and the view being coded. We have defined projection matrices as a multiplication of intrinsic and extrinsic camera parameters that need to be transmitted for each view being coded in the bitstream.

$$\boldsymbol{P} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{R} & -\boldsymbol{R} \cdot \boldsymbol{T} \\ \boldsymbol{O}^T & 1 \end{bmatrix} \tag{4}$$

Necessary distance z of the considered point from camera of interest can be calculated based on the depth sample value v:

$$z = \left(\frac{v}{2^{DepthMapBitDepth}} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}} \right)^{-1} \tag{5}$$

where Z_{near} and Z_{far} are depth maps normalization parameters transmitted in the bitstream, and $DepthMapBitDepth$ is bit depth of the depth sample value (typically 8 or 16 bits per sample is used).

Actual camera parameters for each view are a natural extension of the simplified camera parameters already used in 3D-HEVC. Beside already transmitted parameters: focal length

f_x, optical center c_x of the camera and position of the camera optical center along x axis T_x, we have to transmit additional intrinsic parameters: second focal length along vertical direction f_y, position of the optical center along second axis c_y, and extrinsic parameters: rotation matrix $\boldsymbol{R}$ and remaining coordinates of camera position T_y and T_z. T_y and T_z together with T_x create vector $\boldsymbol{T}$.

In the proposed 3D-HEVC extension we have modified disparity vector derivation process in such tools as Disparity Compensated Prediction (DCP), Neighboring Block Disparity Vector (NBDV), Depth oriented NBDV (DoNBDV), View Synthesis Prediction (VSP), Inter-view Motion Prediction (IvMP), Illumination Compensation (IC).

In the 3D-HEVC, contrary to the proposed extension, derived disparity vectors do not depend on the position in view being coded. For example, in the DoNBDV tool, the disparity for a given block is set to the value that corresponds with the maximum value of four corner depth samples value of virtual depth map block. In the proposed method, the disparity is calculated based on half of the maximum depth sample value and the position of selected corner of the block.

Considering circular (arc) view arrangements, a given depth sample value can represent a different distance plane for each view (which is not the case in 1D parallel view arrangement). In the proposed extension, corresponding depth sample value in the reference view can be derived simultaneously with the disparity vector through the equation (3). The equation (3) provides distance z_r of the correspondent point from the reference camera which can be normalized to depth sample value with the help of

$$v = \frac{\frac{1}{z} - \frac{1}{z_{far}}}{\frac{1}{z_{near}} - \frac{1}{z_{far}}} \cdot 2^{DepthMapBitDepth} \tag{6}$$

The derivation of depth sample value has been applied to the synthesis of dependent views' depth maps.

3. EXPERIMENTS

Extensive experiments have been performed in order to evaluate the performance of proposed extension. Our improvements have been implemented in reference 3D-HEVC codec version 13 (HTM 13.0) [3]. For comparison purposes, unmodified 3D-HEVC codec and MV-HEVC codec have been used (also HTM 13.0).

All tests have been performed according to JCT-3V Common Test Conditions (CTC) [4], that describe default encoder configuration for evaluation of 3D codec performance used by International Organization for Standardization. Essential parameters of the encoders have been gathered in Table 1.

Table 1. Essential configuration parameters of the encoders used in the experiments.

Parameter	Value
Profile	Main
GOP size	8
Intra period	24
Hierarchical GOP	Yes
Number of referenced frames	4
Rate-distortion optimization	On
Search range for motion and disparity estimation	±64
Neighboring Block Disparity Vector (NBDV)	On
Depth oriented NBDV (DoNBDV)	On
View Synthesis Prediction (VSP)	On
Inter-view Motion Prediction (IvMP)	On
Illumination Compensation (IC)	On
View synthesis optimization for depth coding (VSO)	Off

As the 3D-HEVC is designed to support only linear camera arrangement, in the case of the encoding sequences with circular (arc) camera arrangement, positions of the cameras were set as distances from the base view. This way, a rough approximation of proper scaling of the disparity vector can be done by the encoder.

We have evaluated the proposed extension with three commonly recognized multiview test sequences with the arc arrangement of the cameras, namely: *Poznan Blocks* [5], *Ballet* and *Break Dancers* [6]. All sequences were provided with high quality depth data for all of the views.

For compatibility purposes, we have also tested our codec against sequences with linear arrangement of the cameras. For that purpose we have used six multiview sequences with depth maps that are recommended by the MPEG for 3D video codec evaluations (*Poznan Street*, *Poznan Hall 2* [7], *Dancer* [8], *Balloons*, *Kendo* [9], *Newspaper* [10]).

Exemplary images from used test sequences have been depicted in figure 1.

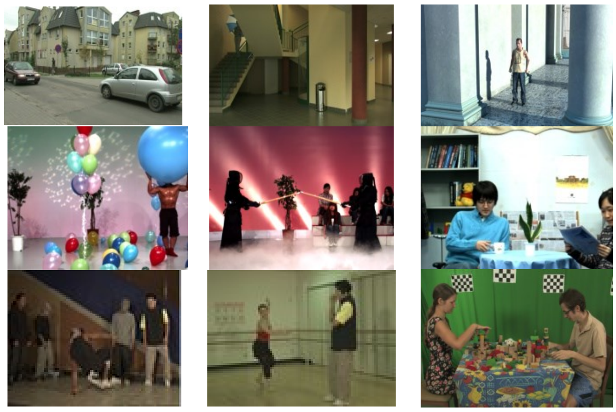

Figure 1. Exemplary pictures from multiview test sequences. From top-left: Poznan Street, Poznan Hall 2, Dancer, Baloons, Kendo, Newspaper, Breakdancer, Ballet, Poznan Blocks.

Three views and three corresponding depth maps of each sequence (see Table 2) have been encoded with proposed codec, original unmodified 3D-HEVC and MV-HEVC at 4 QP values (25, 30, 35, 40). During tests, the total bitrate of all views and all depth maps and average luminance PSNR of all views have been calculated.

Table 2. View positions of each sequences used in the experiment

Sequence	Coded views	View angle
Poznan Street	4, 5, 3	N/A (linear)
Poznan Hall 2	6, 7, 5	N/A (linear)
Dancer	5, 1, 9	N/A (linear)
Balloons	4, 3, 5	N/A (linear)
Kendo	3, 1, 5	N/A (linear)
Newspaper	5, 4, 6	N/A (linear)
Poznan Blocks	5, 4, 6	11 degree
Ballet	4, 3, 5	4 degree
Breakdancers	4, 3, 5	4 degree

4. RESULTS

The averaged luminance PSNR values and the corresponding total bitrates for all test sequences for three analyzed codecs have been gathered and presented in Table 3. RD curves for exemplary sequences with linear (Poznan Street) and arc (Poznan Blocks) camera arrangements have been shown in figures 1 and 2 respectively.

Table 3. Experimental results for all test sequences for analyzed codecs.

Sequence	QP	MV-HEVC		3D-HEVC		Proposed	
		Bitrate [kbit/s]	PSNR [dB]	Bitrate [kbit/s]	PSNR [dB]	Bitrate [kbit/s]	PSNR [dB]

Dancer	25	1389	41.88	1329	41.88	1329	41.88
	30	587	40.90	549	40.91	549	40.91
	35	307	39.48	281	39.52	281	39.52
	40	173	37.69	154	37.77	154	37.77
Poznan Hall2	25	3970	39.24	3849	39.25	3849	39.25
	30	1422	37.30	1357	37.33	1357	37.33
	35	653	35.33	620	35.37	620	35.37
	40	330	33.29	312	33.34	312	33.34
Balloons	25	6686	38.28	6458	38.30	6458	38.30
	30	2812	35.57	2671	35.60	2671	35.60
	35	1295	33.12	1205	33.16	1205	33.16
	40	617	30.85	558	30.91	558	30.91
Newspa-per	25	1587	42.84	1422	42.88	1422	42.88
	30	807	40.56	710	40.65	710	40.65
	35	452	38.11	394	38.22	394	38.22
	40	264	35.52	229	35.67	230	35.67
Poznan Street	25	1647	42.36	1466	42.42	1466	42.42
	30	855	40.11	754	40.22	754	40.22
	35	481	37.49	423	37.64	423	37.64
	40	277	34.70	246	34.88	246	34.88
Kendo	25	1710	40.49	1614	40.53	1614	40.52
	30	845	38.12	788	38.16	788	38.16
	35	457	35.64	421	35.68	420	35.68
	40	260	33.12	238	33.18	238	33.18
Poznan Blocks	25	2761	43.41	2736	43.34	2655	43.42
	30	1520	40.82	1509	40.74	1446	40.84
	35	868	38.05	859	37.94	813	38.05
	40	496	35.18	491	35.08	458	35.21
Ballet	25	936	41.62	892	41.58	871	41.63
	30	476	40.12	450	40.04	438	40.12
	35	267	38.18	250	38.07	242	38.18
	40	152	35.89	143	35.82	138	35.93
Break Dancers	25	2352	39.17	2216	39.13	2162	39.16
	30	964	37.71	890	37.67	860	37.72
	35	491	36.02	457	35.98	437	36.05
	40	261	34.10	246	34.08	233	34.15

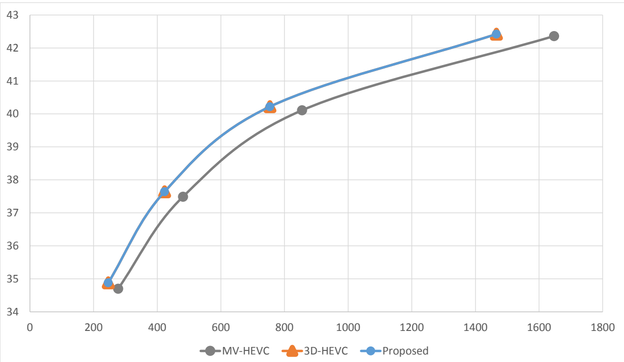

Figure 1. RD curves for Poznan Street sequence (linear camera arrangement) for MV-HEVC, 3D-HEVC and proposed codecs.

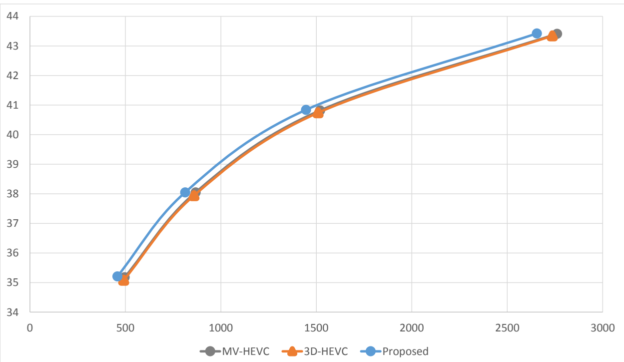

Figure 2. RD curves for Poznan Blocks sequence (arc camera arrangement) for MV-HEVC, 3D-HEVC and proposed codecs.

Table 4. Average bitrate changes calculated as Bjøntegaards rates for PSNR Y introduced by the proposed codec against 3D-HEVC and MV-HEVC for sequences with linear and arc view arrangements.

	Sequence	BD-Rate [%]		
View arrange-ment		Proposed vs 3D-HEVC	Proposed vs MV-HEVC	3D-HEVC vs MV-HEVC
Linear	Poznan Street	0,05	-7.10	-7.11
	Poznan Hall 2	0,09	-9.14	-9.18
	Dancer	0,02	-14.34	-14.36
	Balloons	0,08	-8.21	-8.20
	Kendo	0,09	-6.05	-6.08
	Newspaper	0,04	-14.26	-14.28
	Average	**0,06**	**-9.85**	**-9.87**
Arc	Poznan Blocks	-6,17	-11.11	-5.28
	Ballet	-5,68	-8.82	-3.35
	Break Dancers	-6,89	-5.94	1.02
	Average	**-6,25**	**-8.62**	**-2.54**

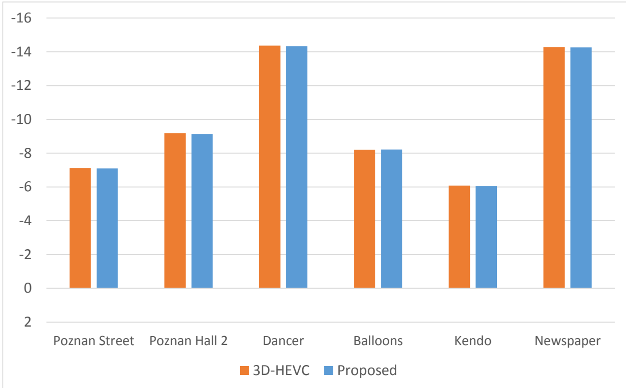

Figure 3. Bitrate reduction (Bjøntegaards rates for PSNR Y) against MV-HEVC for sequences with linear view arrangement

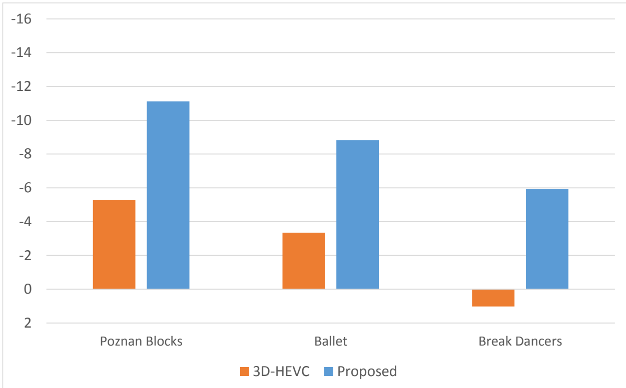

Figure 4. Bitrate reduction (Bjøntegaards rates for PSNR Y) against MV-HEVC for sequences with arc view arrangement

In order to assess the coding performance of the proposed codec, commonly used Bjøntegaard delta has been calculated [11] based on the data presented in Table 3. Achieved results have been gathered in Table 4 and depicted in figures 3 and 4. The sequences have been divided by the view arrangement (linear and arc).

For circular (arc) camera setup, our improved codec allows for about 6% bitrate reduction in comparison with 3D-HEVC and over 8% bitrate reduction in the case of MV-HEVC.

For linear (parallel) camera setup for which 3D-HEVC was designed, proposed modifications caused negligibly small bitrate increase while preserving the same quality of encoded video – on average 0.06%. Due to the fact that the proposed codec has been built on the top of the 3D-HEVC, the achieved bitstream reduction against MV-HEVC is almost the same as for 3D-HEVC against MV-HEVC (about 9.9%).

The small performance drop for sequences with linear view arrangement is caused mainly by the necessity of transmission of full camera parameters while in original 3D-HEVC only simplified version of camera parameters is used and transmitted.

Therefore, our modifications do not only deteriorate the quality of encoded video for linear camera setting, but also lead to a significant improvement in compression efficiency for sequences with arc arranged cameras.

5. CONCLUSIONS

In the paper, modifications of the 3D-HEVC codec, that allow to utilize knowledge of the arrangement of camera setup (in our case – arc camera setup), have been presented. The performance of the proposed new codec has been compared to the state-of-the-art codec 3D-HEVC (HTM version 13.0) and MV-HEVC. The performed experiments showed that the proposed modifications led on average to over 6% bitrate reduction for video sequences recorded with the arc arranged camera setup comparing to the unmodified 3D-HEVC codec. Whereas, for video sequences with linear camera setup, the observed bitrate change is negligibly small.

The obtained results show that there is still room for further improvement in compression efficiency of the 3D-HEVC standard, especially by relaxing explicit constraints on camera setup.

6. ACKNOWLEDGEMENT

Research project was supported by National Science Centre, Poland, according to the decision DEC 2012/05/B/ST7/01279.

7. REFERENCES

[1] G.J. Sullivan,, J.M. Boyce, Y. Chen, J.-R. Ohm, C.A. Segall, A. Vetro, "Standardized Extensions of High Efficiency Video Coding (HEVC)", *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, Dec 2013.

[2] Müller K., Merkle P., and Wiegand T., "3D Video Representation Using Depth Maps", *Proceedings of the IEEE*, vol. 99, no. 4, pp. 643–656, Apr. 2011.

[3] 3D HEVC reference codec available online https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-13.0.

[4] K. Müller, A. Vetro, "Common Test Conditions of 3DV Core Experiments" *Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 7th Meeting: Doc. JCT3V-G1100*, San José, US, Jan. 2014.

[5] M. Domański, A. Dziembowski, A. Kuehn, M. Kurc, A. Łuczak, D. Mieloch, J. Siast, O. Stankiewicz, K. Wegner, „Poznan Blocks – a multiview video test sequence and camera parameters for Free Viewpoint Television" *ISO/IEC JTC1/SC29/WG11 Doc. M32243*, San Jose, USA, Jan. 2014.

[6] C.L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, "High-quality video view interpolation using a layered representation," ACM Transactions on Graphics, vol. 23, no. 3. pp. 600-608, Aug. 2004.

[7] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, "Poznań multiview video test sequences and camera parameters," *ISO/IEC JTC1/SC29/WG11 MPEG Doc. M17050*, Xian, China, Oct. 2009.

[8] D. Rusanovskyy, P. Aflaki, M.M. Hannuksela, "Undo Dancer 3DV sequence for purposes of 3DV standardization," *ISO/IEC JTC1/SC29/WG11 MPEG Doc. M20028*, Geneva, Switzerland, Mar. 2011.

[9] M. Tanimoto, T. Fujii, N. Fukushima, "1D parallel test sequences for MPEG-FTV," *ISO/IECJTC1/SC29/WG11 MPEG Doc. M15378*, Archamps, France, Apr. 2008.

[10] Y.S. Ho, E.K. Lee, C. Lee, "Multiview video test sequence and camera parameters," *ISO/IECJTC1/SC29/WG11 MPEG Doc. M15419*, Archamps, France, Apr. 2008.

[11] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," *ITU-T SG16, Doc. VCEG-M33,* Austin, USA, Apr. 2001.