INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG04
MPEG VIDEO CODING

ISO/IEC JTC 1/SC 29/WG 04 m59516
April 2022, Online

Title: Decoder-side depth estimation with extended input depth assistance
Source: PUT: Dawid Mieloch, Adrian Dziembowski, Błażej Szydełko, Dominika Klóska
ETRI: Gwangsoon Lee, Jun Young Jeong

Abstract

This document presents a description of the extension of the MIV DSDE, where we send full depth maps for basic views from the 0th attribute atlas, and depth patches for basic views from 1st and 2nd attribute atlases. This depth information helps the IVDE to obtain better quality and significantly decrease the computational time of depth estimation.

1 Proposed approach

The proposal is an extension of the approach presented in m58048, which was the combination of V17 and G17 configuration, and the geometry was sent only for views packed into the 0th of three texture atlases.

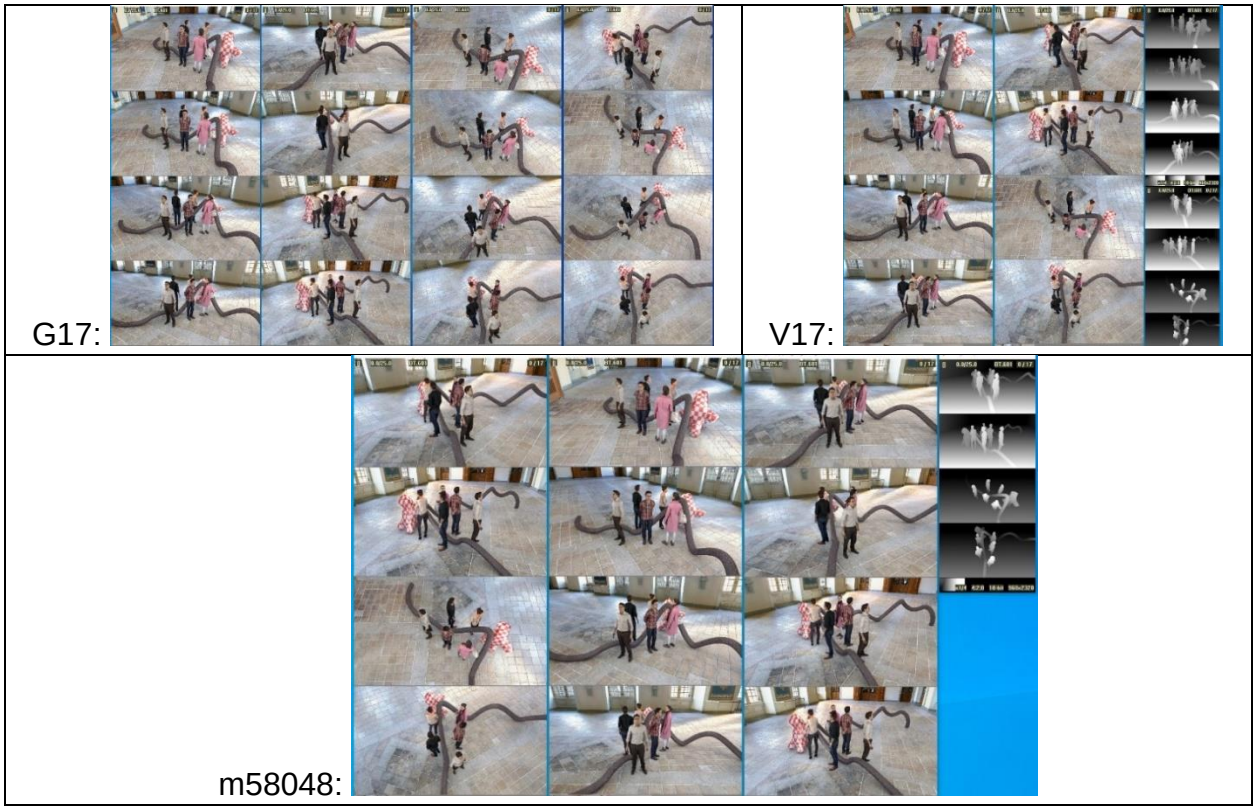

Fig. 1. Atlases in different approaches: G17, V17, and m58048.

In m58048, for basic views from the 1st, and 2nd atlas, the geometry information was not being sent at all. For these views, depth maps were estimated at the decoder side, basing on textures and depth maps already available in the decoder.

In the proposed extended approach, we send the depth information for views from all three attribute atlases:

- full depth maps for views from the 0th attribute atlas (highlighted in yellow in Fig. 2),
- depth patches (containing non-pruned pixels) for views from 1st and 2nd atlas (red highlight in Fig. 2).

Fig. 2. Atlases in proposed extended approach.

1.1 Which regions should be sent within depth patches?

In the proposed approach, we want to provide IVDE input depth maps for all transmitted input views. For views from the 0th atlas, such depth maps are being transmitted, but for views from atlases 1 and 2, they have to be rendered by the MIV decoder.

Such a rendering reprojects pixels from sent depth maps into the position of remaining views, allowing to estimate all the required depth maps. However, as presented in Fig. 3C, when no additional depth information is used, some regions of the rendered view are rendered incorrectly (background pixels instead of pixels representing the foreground knight).

When using the depth map for the rendered view from Fig. 3C as input for depth estimation, the IVDE will not be able to properly estimate depth in this region (it will have input depth value, so it will not try to change it into the depth of the knight).

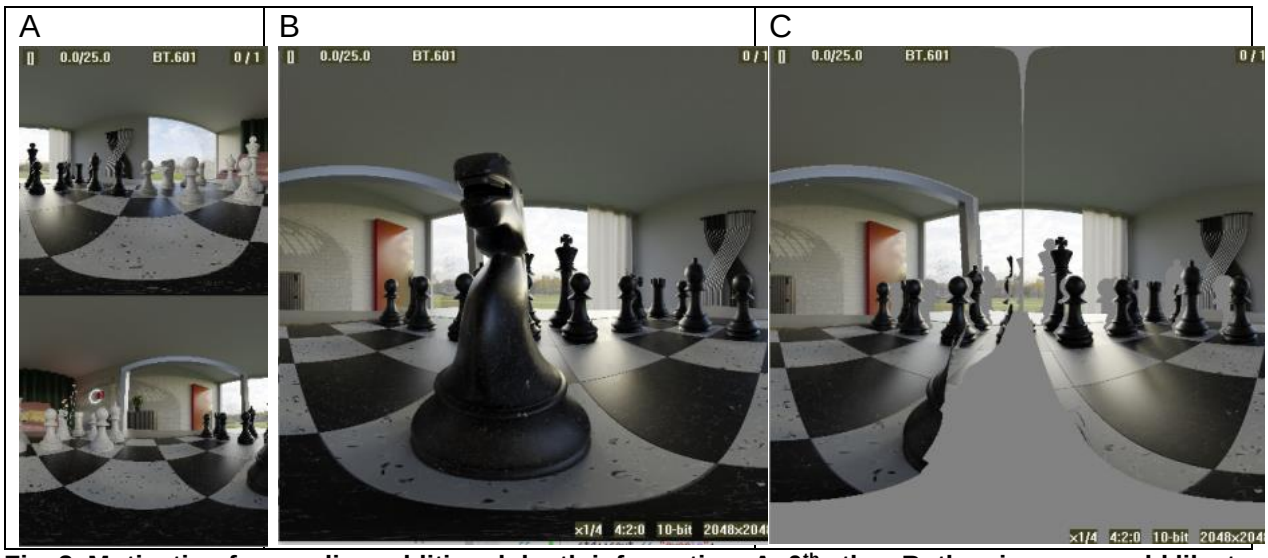

Fig. 3. Motivation for sending additional depth information. A: 0th atlas, B: the view we would like to render in the MIV decoder, C: view rendered using information from 0th atlas.

Therefore, for all the regions from views from 1st and 2nd atlases, we have to preserve only pixels, which are closer than pixels reprojected from basic views (and additional views higher in the pruning graph).

1.2 MIV encoding

In order to provide good quality depth information for the large part of the scene, the basic views are reshuffled (as in m58048), and the first atlas contains the most distant views (chosen by the TMIV view selector/labeler launched for the 2nd time, only for basic views).

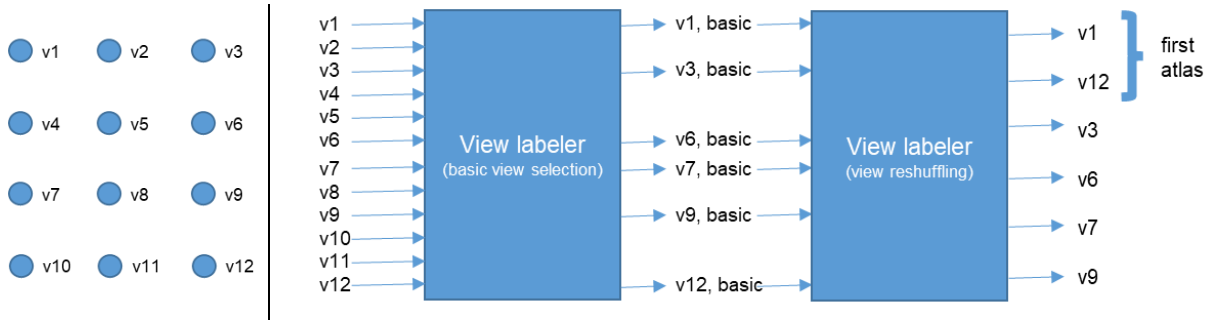

Fig. 4. Basic view reshuffling in case, where an atlas contains 2 views. Left: camera arrangement, right: view selection process (basic views are additionally processed to find the most distant ones).

The proposal utilizes syntax elements already available in the MIV Extended profile:

- `vps_geometry_video_present_flag[atlasID]` is set to 0 for `atlasID` == 1 and 2,
- `vps_attribute_video_present_flag[atlasID]` is set to 0 for `atlasID` == 3.

To allow pruning of depth for some views while preserving the entire texture, the viewParamsList in the TMIV encoder contains:

- views from 0th atlas (as basic views),
- views from 1st and 2nd atlas (as basic views),
- views from 1st and 2nd atlas once again (as additional views).

Only the third group of views is being pruned, and the texture for these views is omitted in transmission. Views from the first two groups are pasted into atlases without pruning. For views from the 2nd group, no depth is being sent.

In total, there are 5 atlases instead of 4. However, geometry atlases are much smaller than the constrain, so they could be packed into one video stream.

1.3 MIV decoding

On the decoder side, the input views and corresponding depth maps have to be restored before the depth estimation step. The views are just reconstructed in the unpacking process, while the corresponding depth maps (input depth maps for IVDE) are rendered.

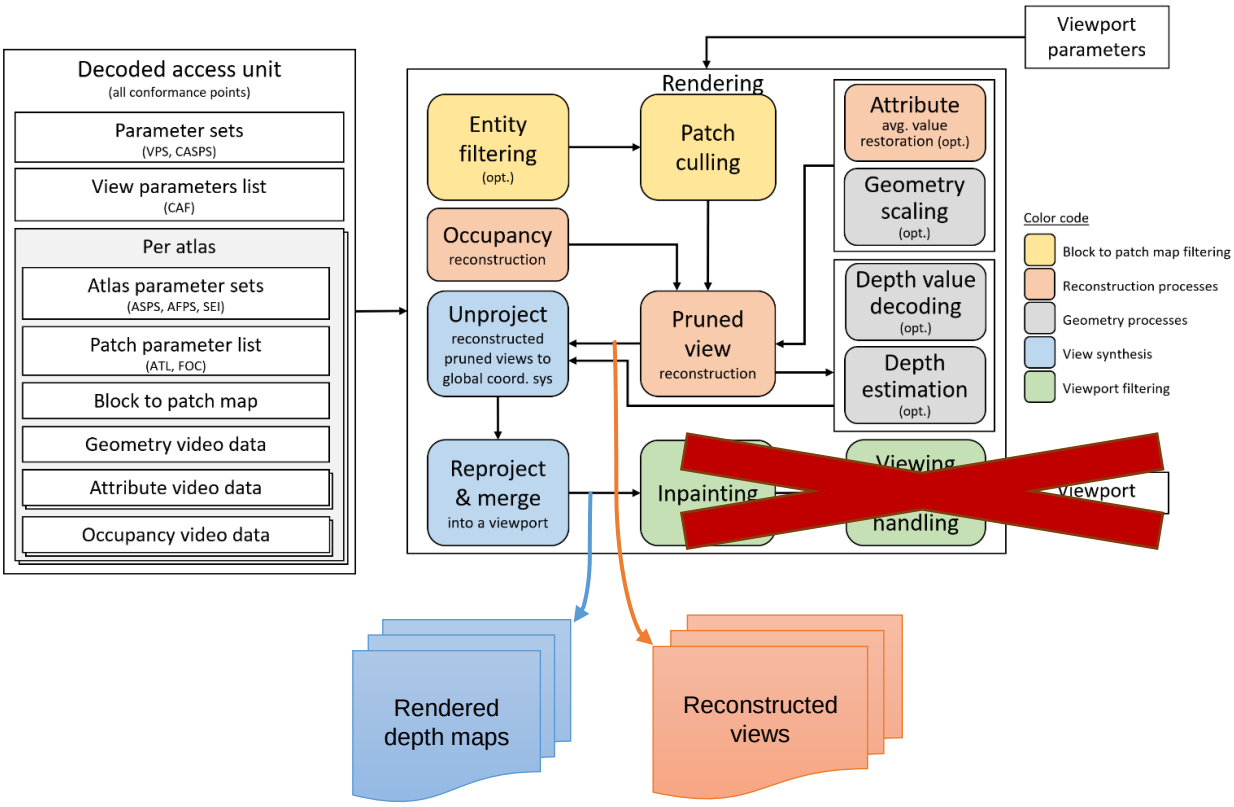

Fig. 5. MIV decoding before the depth estimation.

1.4 Depth estimation

The IVDE receives a set of input views and a depth map for each. For views from the 0th atlas, the input depth maps are fully occupied. Depth maps for other views may contain holes, which will be filled by the depth estimation process (Fig. 6).

Fig. 6. Input depth maps (left) vs. output depth maps (right); fully occupied depth maps for views from 0th atlas (1st and 2nd rows), and depth maps with holes for views from atlases 1 and 2 (3rd and 4th rows).

1.5 MIV rendering

The rendering of final viewports is performed in the exactly same way, as in TMIV11. In the experiment, we used TMIV11 renderer without any changes.

2 Results

G17 anchor vs. the proposal with standard depth QP = $\max(1, [-14.2 + 0.8q])$

Mandatory content - Proposal vs. Low/High-bitrate Anchors																	
								Runtime ratio (%)			Max delta Y-PSNR [dB]			Max delta IV-PSNR [dB]		
Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Pixel rate [%]	Pixel rate [GP/s]	Frame rate [Hz]	Atlas encoding	Video encoding	Decoding & Rendering	MIV DSDE	#######	Difference [%]	MIV DSDE	#######	Difference [%]
ClassroomVideo	A	-91.1%	-55.3%	-58.6%	-35.6%	######	#####	30	1352.3%	89.8%	25.5%	5.69	3.75	-34.2%	4.05	2.41	-40.6%
Museum	B	-59.7%	-42.7%	-35.3%	-25.7%	######	#####	30	2004.3%	103.9%	55.4%	9.18	6.69	-27.1%	6.34	4.46	-29.6%
Fan	O	91.1%	105.2%	68.0%	90.0%	######	#####	30	#######	102.3%	33.7%	10.89	10.08	-7.4%	10.03	9.18	-8.5%
Kitchen	J	-27.6%	-19.2%	-25.6%	-17.1%	######	#####	30	5416.4%	97.0%	39.2%	11.99	11.04	-8.0%	11.21	9.78	-12.7%
Painter	D	51.3%	99.9%	32.0%	79.0%	######	#####	30	6109.9%	83.8%	31.4%	7.60	5.68	-25.3%	7.35	3.23	-56.1%
Frog	E	30.1%	37.2%	35.4%	40.4%	######	#####	30	5566.8%	103.2%	11.1%	7.40	7.12	-3.8%	7.17	7.39	3.1%
Carpark	P	117.5%	110.9%	61.0%	71.9%	######	#####	25	1948.3%	96.3%	36.7%	10.24	9.86	-3.7%	8.19	7.65	-6.7%
Chess	N	279.9%	392.0%	74.6%	35.5%	######	#####	30	2929.6%	108.7%	63.0%	25.19	23.29	-7.6%	23.89	22.98	-3.8%
Group	R	---	---	---	---	######	#####	30	#######	98.7%	28.6%	22.60	18.70	-17.3%	23.55	17.47	-25.8%
MIV		---	---	---	---	######	#####		**6101.3%**	**98.2%**	**36.1%**	**12.31**	**10.69**	**-14.9%**	**11.31**	**9.39**	**-20.1%**
Optional content - Proposal vs. Low/High-bitrate Anchors																	
Fencing	L	155.2%	94.7%	87.7%	91.0%	######	#####	25	3181.0%	84.9%	32.8%	12.90	13.31	3.3%	9.18	9.79	6.6%
Hall	T	67.0%	62.2%	-37.3%	-11.5%	######	#####	25	3792.3%	81.7%	44.8%	16.13	16.03	-0.6%	13.57	13.01	-4.1%
Street	U	27.2%	33.3%	30.5%	38.1%	######	#####	25	2352.3%	97.9%	47.4%	7.07	6.68	-5.5%	4.91	4.42	-9.9%
ChessPieces	Q	---	---	---	---	######	#####	30	2448.0%	111.2%	68.4%	27.71	29.07	4.9%	25.79	27.99	8.5%
Hijack	C	---	---	---	---	######	#####	30	1195.7%	111.4%	45.7%	22.33	18.91	-15.3%	21.03	17.50	-16.8%
Mirror	I	64.8%	74.2%	66.7%	71.5%	######	#####	30	7311.6%	94.8%	40.7%	12.41	11.74	-5.4%	11.17	11.09	-0.8%
Cadillac	G	46.4%	55.7%	8.4%	23.2%	######	#####	30	7794.1%	117.3%	26.1%	14.30	12.63	-11.6%	14.29	13.11	-8.2%
MIV		---	---	---	---	######	#####		**4010.7%**	**99.9%**	**43.7%**	**16.12**	**15.48**	**-4.3%**	**14.28**	**13.84**	**-3.5%**

G17 anchor vs. the proposal with modified depth QP = $\max(1, 0.8q)$

Mandatory content - Proposal vs. Low/High-bitrate Anchors																	
								Runtime ratio (%)			Max delta Y-PSNR [dB]			Max delta IV-PSNR [dB]		
Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Pixel rate [%]	Pixel rate [GP/s]	Frame rate [Hz]	Atlas encoding	Video encoding	Decoding & Rendering	MIV DSDE	#######	Difference [%]	MIV DSDE	#######	Difference [%]
ClassroomVideo	A	---	-69.3%	-64.7%	-47.9%	######	#####	30	2185.6%	100.1%	22.0%	5.69	3.76	-33.9%	4.05	2.41	-40.6%
Museum	B	-50.5%	-36.3%	-35.5%	-29.3%	######	#####	30	2023.7%	119.5%	59.6%	9.18	7.07	-23.0%	6.34	4.78	-24.5%
Fan	O	104.4%	65.4%	62.3%	39.3%	######	#####	30	#######	120.9%	37.5%	10.89	11.16	2.4%	10.03	10.43	4.0%
Kitchen	J	-23.1%	-16.0%	-28.7%	-21.7%	######	#####	30	4780.0%	94.7%	41.6%	11.99	10.89	-9.2%	11.21	9.78	-12.7%
Painter	D	-11.9%	2.0%	-24.8%	-8.7%	######	#####	30	5386.2%	104.0%	32.9%	7.60	6.16	-19.0%	7.35	3.61	-50.9%
Frog	E	21.2%	19.6%	23.2%	19.9%	######	#####	30	3844.5%	106.8%	12.3%	7.40	7.55	2.0%	7.17	7.75	8.1%
Carpark	P	68.3%	53.2%	20.2%	22.6%	######	#####	25	2048.1%	107.2%	43.0%	10.24	10.29	0.5%	8.19	8.08	-1.4%
Chess	N	---	---	155.4%	92.7%	######	#####	30	3936.2%	105.2%	70.8%	25.19	23.93	-5.0%	23.89	23.23	-2.8%
Group	R	---	---	---	---	######	#####	30	#######	114.0%	39.8%	22.60	18.73	-17.1%	23.55	17.62	-25.2%
MIV		---	---	---	---	######	#####		**6783.5%**	**108.0%**	**39.9%**	**12.31**	**11.06**	**-11.4%**	**11.31**	**9.74**	**-16.2%**
Optional content - Proposal vs. Low/High-bitrate Anchors																	
Fencing	L	102.9%	53.5%	54.4%	51.2%	######	#####	25	3340.4%	96.0%	35.3%	12.90	13.58	5.3%	9.18	10.09	10.0%
Hall	T	-34.3%	-37.7%	-81.5%	-78.7%	######	#####	25	2571.5%	101.6%	45.8%	16.13	16.28	0.9%	13.57	13.31	-1.9%
Street	U	-3.4%	-2.7%	-0.8%	-1.6%	######	#####	25	2956.8%	124.1%	42.3%	7.07	7.01	-0.9%	4.91	4.69	-4.3%
ChessPieces	Q	---	---	---	---	######	#####	30	2583.9%	110.9%	70.0%	27.71	29.18	5.3%	25.79	28.14	9.1%
Hijack	C	---	---	---	---	######	#####	30	1258.9%	124.6%	47.3%	22.33	18.99	-15.0%	21.03	17.60	-16.3%
Mirror	I	38.9%	37.4%	41.2%	32.6%	######	#####	30	9639.9%	112.3%	42.3%	12.41	12.13	-2.2%	11.17	11.46	2.6%
Cadillac	G	109.3%	73.8%	-2.0%	2.7%	######	#####	30	#######	129.3%	28.3%	14.30	12.82	-10.3%	14.29	13.08	-8.4%
MIV		---	---	---	---	######	#####		**4854.9%**	**114.1%**	**44.5%**	**16.12**	**15.71**	**-2.4%**	**14.28**	**14.05**	**-1.3%**

The modification of depth QP significantly decreased bitrate required for geometry atlases, therefore, the quality for texture is not much lower than for the G17 anchor. On the other hand, the quality of highly compressed depth maps is improved by IVDE (Figs. 7 and 8).

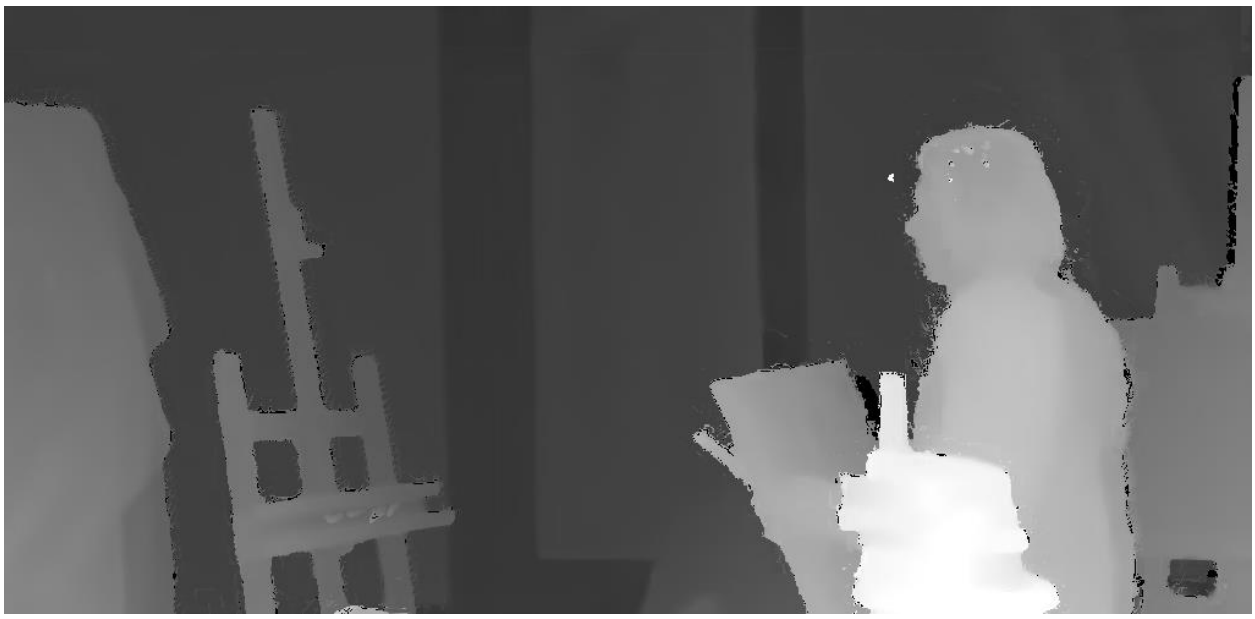

Fig. 7. Input depth map for Painter sequence at QP5.

Fig. 8. Output depth map for Painter sequence at QP5.

The side-by-side posetraces (G17 vs the proposal) were uploaded to MPEG-I/Part12-ImmersiveVideo/for_testing/m59616_depth_assistance. This comparison shows significant improvement of the quality over the traditional DSDE. The biggest differences can be found in SA p01, SB p02, SC p02, SD p01, SG p03, SN p01, SQ p02, SR p03.

A17 anchor vs. the proposal with modified depth QP = $\max(1, 0.8q)$

Mandatory content - Proposal vs. Low/High-bitrate Anchors

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Pixel rate [%]	Pixel rate [GP/s]	Frame rate [Hz]	Runtime ratio (%) Atlas encoding	Video encoding	Decoding & Rendering	Max delta Y-PSNR [dB] MIV view	#######	Difference [%]	Max delta IV-PSNR [dB] MIV view	#######	Difference [%]
ClassroomVideo	A	97.3%	-17.1%	14.1%	-26.1%	######	#####	30	10.2%	151.4%	247.3%	0.99	3.76	281.2%	0.76	2.41	218.6%
Museum	B	77.5%	-1.9%	34.3%	-12.0%	######	#####	30	8.1%	96.7%	1134.9%	9.45	7.07	-25.2%	5.35	4.78	-10.7%
Fan	O	---	---	-60.5%	-73.0%	######	#####	30	133.0%	202.7%	2995.4%	8.02	11.16	39.2%	7.24	10.43	44.1%
Kitchen	J	-39.7%	-32.0%	14.9%	7.3%	######	#####	30	35.4%	106.2%	2449.0%	14.67	10.89	-25.8%	11.19	9.78	-12.7%
Painter	D	-62.9%	-56.4%	-46.4%	-48.1%	######	#####	30	70.4%	114.4%	3177.6%	7.94	6.16	-22.4%	5.26	3.61	-31.3%
Frog	E	-59.2%	-50.8%	-38.2%	-40.9%	######	#####	30	65.5%	124.9%	2339.1%	7.39	7.55	2.1%	7.21	7.75	7.5%
Carpark	P	14.0%	-7.3%	19.2%	-7.8%	######	#####	25	37.8%	114.3%	1828.3%	7.05	10.29	45.8%	5.01	8.08	61.2%
Chess	N	---	---	---	---	######	#####	30	35.1%	97.4%	2470.0%	13.60	23.93	76.0%	12.44	23.23	86.7%
Group	R	118.2%	50.8%	174.7%	54.2%	######	#####	30	147.2%	129.4%	2647.6%	12.89	18.73	45.3%	10.30	17.62	71.1%
MIV		---	---	---	---	######	#####		**60.3%**	**126.4%**	**2143.2%**	**9.11**	**11.06**	**46.2%**	**7.20**	**9.74**	**48.3%**

Optional content - Proposal vs. Low/High-bitrate Anchors

Fencing	L	76.7%	-15.4%	42.0%	-25.3%	######	#####	25	46.6%	115.9%	1901.4%	10.37	13.58	31.0%	7.60	10.09	32.8%
Hall	T	---	31.9%	63.4%	-16.8%	######	#####	25	38.0%	157.8%	1702.8%	11.67	16.28	39.6%	8.27	13.31	61.1%
Street	U	-64.4%	-47.3%	-30.1%	-28.1%	######	#####	25	56.3%	124.5%	1600.0%	8.48	7.01	-17.4%	4.54	4.69	3.3%
ChessPieces	Q	---	---	---	---	######	#####	30	18.6%	79.0%	2455.6%	14.44	29.18	102.2%	15.29	28.14	84.1%
Hijack	C	---	---	---	112.1%	######	#####	30	8.8%	134.3%	388.6%	7.98	18.99	137.9%	5.70	17.60	208.7%
Mirror	I	-2.6%	-26.1%	25.8%	-19.1%	######	#####	30	131.6%	114.9%	2199.2%	8.76	12.13	38.5%	5.23	11.46	119.1%
Cadillac	G	-45.0%	-56.3%	-39.1%	-54.0%	######	#####	30	118.1%	73.3%	2444.3%	12.08	12.82	6.1%	11.16	13.08	17.2%
MIV		---	---	---	---	######	#####		**59.7%**	**114.2%**	**1813.1%**	**10.54**	**15.71**	**48.3%**	**8.26**	**14.05**	**75.2%**

3 Recommendation

We recommend opening the Exploration Experiment which will test the proposal with different configurations (tuning of depth QP, IVDE parameters).

4 Acknowledgement

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00207, Immersive Media Research Laboratory).