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

ISO/IEC JTC1/SC29/WG11 MPEG2020/M54277
June 2020, Online

Source	**Poznań University of Technology (PUT), Poznań, Poland**
Status	**Input**
Title	**EE Depth results**
Author	Dawid Mieloch, Adrian Dziembowski

1 Introduction

This document presents the comparison of DERS and IVDE prepared using the current EE Depth Common Test Conditions and using TMIV 5.

2 Experimental results

2.1 EE Depth CTC experiment

In the first experiment, DERS [N19143] and IVDE [N19224] were compared using the current Common Test Conditions for EE Depth [N19221].

Few minor changes were made in comparison with EE Depth CTC:

- unified Znear and Zfar for all views of SD (1.8 and 4.4, also used in previous EE Depth),
- changed Znear for SL and SG, as some parts of the scene are closer and do not fit in the previous range (see Fig. 1 and Fig. 2).

Both compared methods used their unified sets of parameters [M53407], [M53527]. This document includes a package with scripts that can be used to generate presented results.

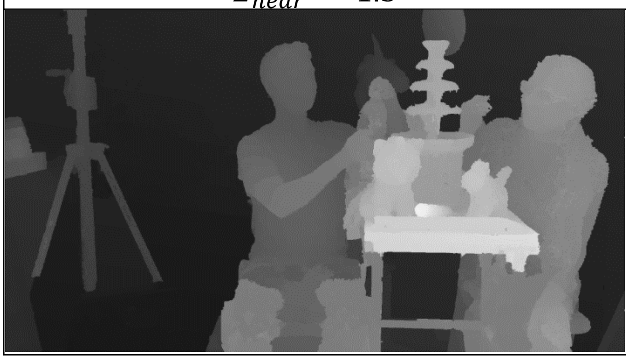
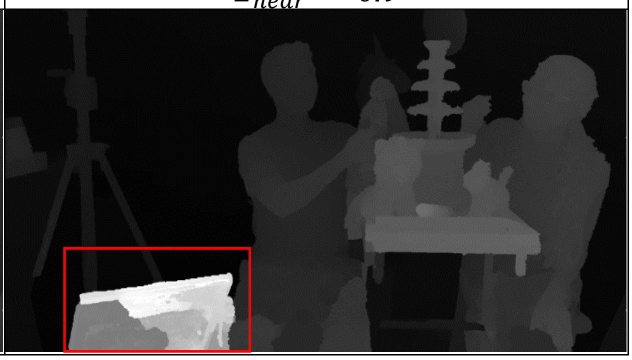
$Z_{near} = 1.5$	$Z_{near} = 0.9$
	

Fig 1. Comparison of SG depth maps with different Z near values.

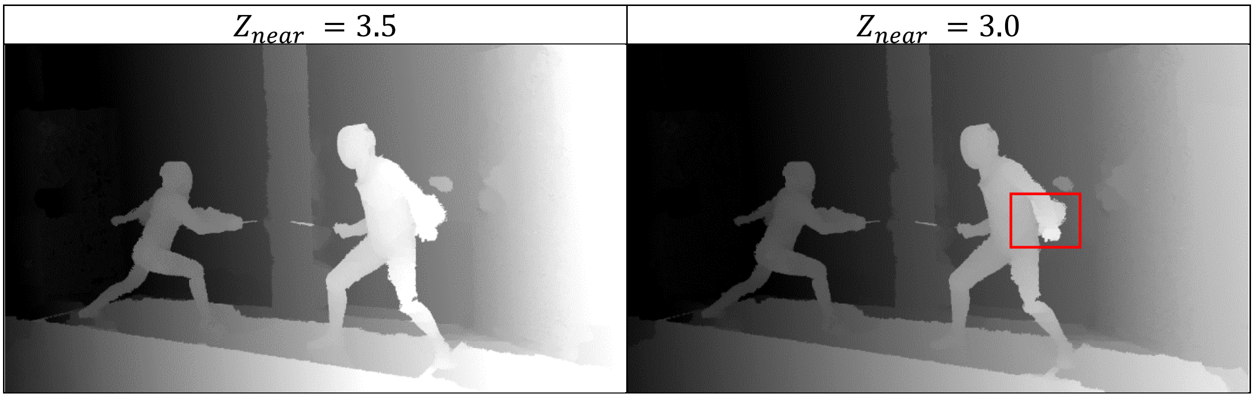


Fig 2. Comparison of SL depth maps with different Z near values.

The results for mandatory sequences are presented in Table 1 (green color – IVDE is better). On average, IVDE achieves higher quality for WS-PSNR-U and V, and for IV-PSNR. IVDE estimates depth for segmented input views, therefore, in point-to-point comparison to input view, the synthesis can be of slightly worse quality. Note that chromas are decimated and IV-PSNR ignores slight shifts of objects in synthesized views.

Max delta of WS-PSNR-Y is on average smaller for IVDE, what is the result of higher inter-view consistency of estimated depth maps. VMAF was shown to provide rather chaotic results, not linked to any other metric. Total computational time is on average almost 3 times shorter for IVDE.

Table 1. DERS and IVDE comparison for mandatory EE Depth sequences.

DERS vs IVDE									
Mandatory sequences									
Sequence	**ID**	**Method**	**Mean WS-PSNR Y**	**Mean WS-PSNR U**	**Mean WS-PSNR V**	**Max delta WS-PSNR Y**	**Mean IV-PSNR**	**Mean VMAF**	**Total computational time**
OrangeShaman	SH	Anchor	34.024978	44.285611	42.053122	6.5281	41.161178	73.260537	117879
		Proposal	33.801844	44.4104	42.157044	5.2405	41.642733	70.581246	35750
		Difference	-0.2231333	0.1247889	0.1039222	-1.2876	0.4815556	-2.6792907	82129
OrangeDancing	SI	Anchor	31.792811	49.458456	51.309	5.1619	42.161144	81.709822	123487
		Proposal	30.392056	48.6013	50.887178	4.7074	40.614678	78.625171	27745
		Difference	-1.4007556	-0.8571556	-0.4218222	-0.4545	-1.5464667	-3.0846512	95742
OrangeKitchen	SJ	Anchor	27.791111	41.674911	45.488211	10.101	34.600133	77.798011	104234
		Proposal	27.923633	41.714178	45.507056	10.4463	35.246256	75.372742	17973
		Difference	0.1325222	0.0392667	0.0188444	0.3453	0.6461222	-2.4252686	86261
TechnicolorPainter	SD	Anchor	32.944363	45.992075	45.06325	6.7205	40.731963	83.968775	61248
		Proposal	31.738125	45.68345	44.92295	5.2703	38.95435	81.687814	32080
		Difference	-1.2062375	-0.308625	-0.1403	-1.4502	-1.7776125	-2.2809607	29168
IntelFrog	SE	Anchor	27.847171	41.580257	40.195014	5.7408	37.603171	79.512381	103412
		Proposal	27.966843	42.099386	40.952514	5.3415	37.678086	80.148127	27020
		Difference	0.1196714	0.5191286	0.7575	-0.3993	0.0749143	0.635746	76392
ULBUnicornA	SF	Anchor	28.071422	41.732633	42.095989	5.7983	37.434611	84.144634	14918
		Proposal	26.654433	41.0926	41.506089	5.6396	36.668711	81.166312	4510
		Difference	-1.4169889	-0.6400333	-0.5899	-0.1587	-0.7659	-2.9783223	10408
ULBBabyUnicorn	SG	Anchor	22.794438	35.755238	34.7304	5.9102	29.409138	54.731111	81312
		Proposal	25.594013	37.173525	36.602425	7.5093	32.484763	65.471355	24679
		Difference	2.799575	1.4182875	1.872025	1.5991	3.075625	10.740244	56633
PoznanFencing	SL	Anchor	31.63478	45.64156	44.76544	4.8289	39.34688	82.249984	111501
		Proposal	29.66846	45.49942	44.08964	3.6084	41.20492	62.819409	44023
		Difference	-1.96632	-0.14214	-0.6758	-1.2205	1.85804	-19.430575	67478
Average (perspective)		Anchor	29.612634	43.265093	43.212553	6.3487125	37.806027	77.171907	89748.875
		Proposal	29.217426	43.284282	43.328112	5.9704125	38.061812	74.484022	26722.5
		Difference	-0.3952083	0.0191897	0.1155587	-0.3783	0.2557847	-2.6878848	63026.375

Table 2. DERS and IVDE comparison for optional EE Depth sequences.

Optional sequences									
ETRIBreaktime	SK	Anchor	27.639013	38.232575	39.1919	10.9667	34.434875	73.717011	80564
		Proposal	27.7271	38.508438	39.462463	10.4994	34.704788	75.148229	50626
		Difference	0.0880875	0.2758625	0.2705625	-0.4673	0.2699125	1.4312179	29938
PoznanCarpark	SP	Anchor	32.4365	45.77674	44.00146	5.7248	41.38788	83.874458	63421
		Proposal	32.22384	45.8293	44.05404	5.3011	41.14222	83.581674	36072
		Difference	-0.21266	0.05256	0.05258	-0.4237	-0.24566	-0.2927846	27349
PoznanHall	ST	Anchor	31.70028	44.00504	45.22198	6.547	39.72162	73.531543	78401
		Proposal	31.24656	43.6198	44.99044	5.0504	38.87892	76.868192	19603
		Difference	-0.45372	-0.38524	-0.23154	-1.4966	-0.8427	3.3366484	58798
PoznanStreet	SU	Anchor	33.5377	46.93718	45.51038	2.9722	43.57772	83.212597	71118
		Proposal	33.3277	46.99328	45.58676	2.4665	43.75354	82.103873	40603
		Difference	-0.21	0.0561	0.07638	-0.5057	0.17582	-1.1087244	30515
Average (perspective)		Anchor	31.328373	43.737884	43.48143	6.552675	39.780524	78.583902	73376
		Proposal	31.1313	43.737704	43.523426	5.82935	39.619867	79.425492	36726
		Difference	-0.1970731	-0.0001794	0.0419956	-0.723325	-0.1606569	0.8415893	36650

2.2 TMIV CTC experiment

In the second experiment, we tested estimated depth maps in TMIV 5 [N19213], following the MIV CTC [N19214]. This experiment was performed to test the inter-view and temporal consistency of depth maps, as the performance of TMIV highly depends on these factors. Note that this experiment uses the same depth maps as the previous experiment.

The results for mandatory sequences are presented in Table 3 (green color – IVDE is better). BR-rate curves for each sequence are presented in Fig. 3. On average, IVDE achieves a much better quality of synthesized views. For all metrics, the average BD-rate decrease is higher than 30%.

Table 3. DERS and IVDE comparison in TMIV for mandatory MIV CTC sequences.

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

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	Max delta Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Pixel rate ratio
OrangeKitchen	SJ	-28.7%	-37.1%	15.80	-31.7%	-41.0%	-30.2%	-33.0%	0.62
TechnicolorPainter	SD	8.6%	9.8%	8.12	6.4%	8.3%	23.5%	20.4%	0.63
IntelFrog	SE	-77.5%	-69.2%	11.75	-75.4%	-68.8%	-55.6%	-59.7%	0.62
PoznanFencing	SL	-69.9%	-68.0%	13.69	-33.3%	-44.9%	-81.4%	-75.3%	0.52
MIV		**-41.9%**	**-41.1%**	**12.34**	**-33.5%**	**-36.6%**	**-35.9%**	**-36.9%**	

The posetraces that compare the use of DERS (left side of videos) and IVDE (right side) depth maps are available in MPEG Content Server in MPEG-I/Poznan/m54277/ directory. For each sequence, 3 available posetraces (P01-P03) were generated, each of them both for high (close to Rate 1) and low quality (close to Rate 5). In some cases, in order to match bitrates, used Rates are not equal in both methods.

In posetraces, for practically all sequences IVDE achieves better subjective quality of synthesized views. Moreover, because of higher inter-view consistency of depth maps, IVDE posetraces do not show sudden changes in the quality during the movement, which can be seen while DERS depth maps are used.

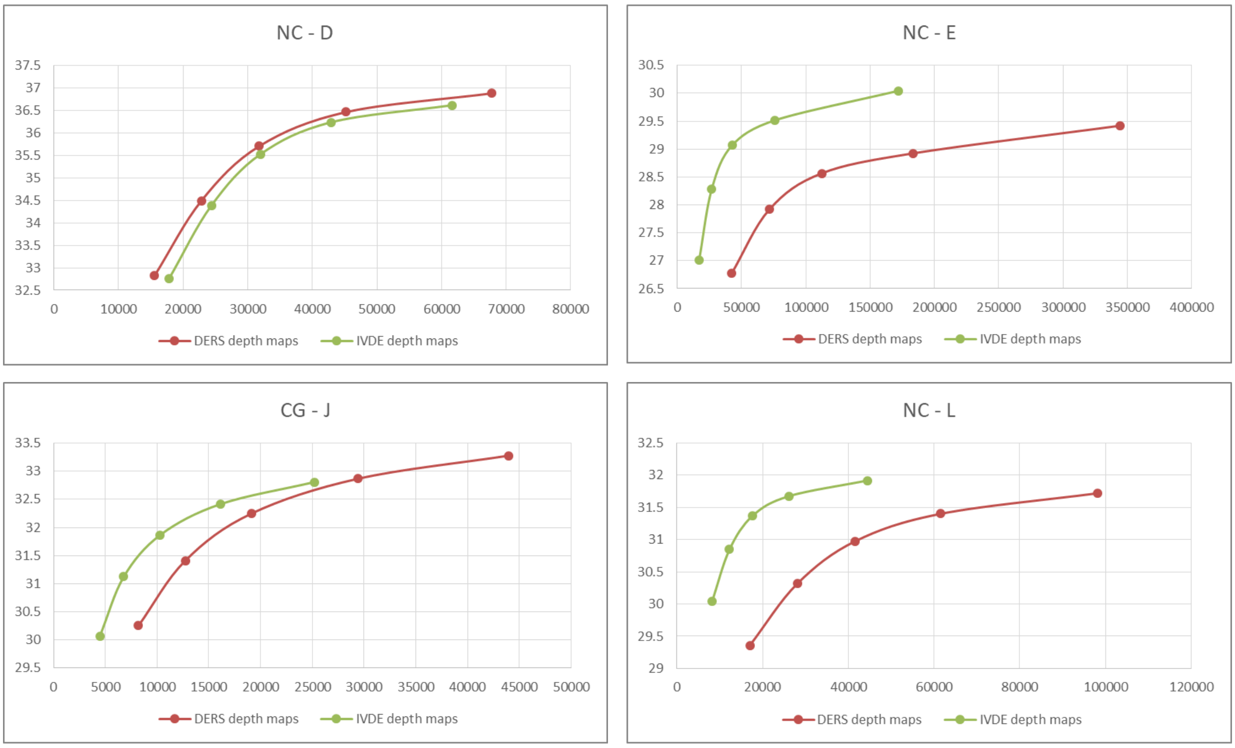


Fig 3. DERS and IVDE comparison in TMIV for mandatory MIV CTC sequences.

The results for optional sequences are presented in Table 4 (green color – IVDE is better). In this case, DERS achieves better average objective quality for PSNR and VMAF, while IV-PSNR is higher for IVDE.

Table 4. DERS and IVDE comparison in TMIV for optional MIV CTC sequences.

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

PoznanCarpark	SP	102.4%	55.0%	12.59	172.7%	70.5%	24.1%	16.5%	0.52
PoznanHall	ST	-29.9%	-34.8%	10.87	-10.2%	-24.2%	-42.1%	-41.8%	0.52
PoznanStreet	SU	160.8%	52.4%	12.41	94.3%	23.9%	-3.8%	-11.2%	0.52
MIV		**77.8%**	**24.2%**	**11.96**	**85.6%**	**23.4%**	**-7.3%**	**-12.2%**	

However, for SP and SU, depth maps generated by DERS are very inconsistent, therefore, much more data have to be put into atlases (Fig. 4 and Fig. 5). It results in sending almost all input views in atlases (especially in SU). In posetraces, IVDE achieves again better, much more inter-view consistent, objective quality.

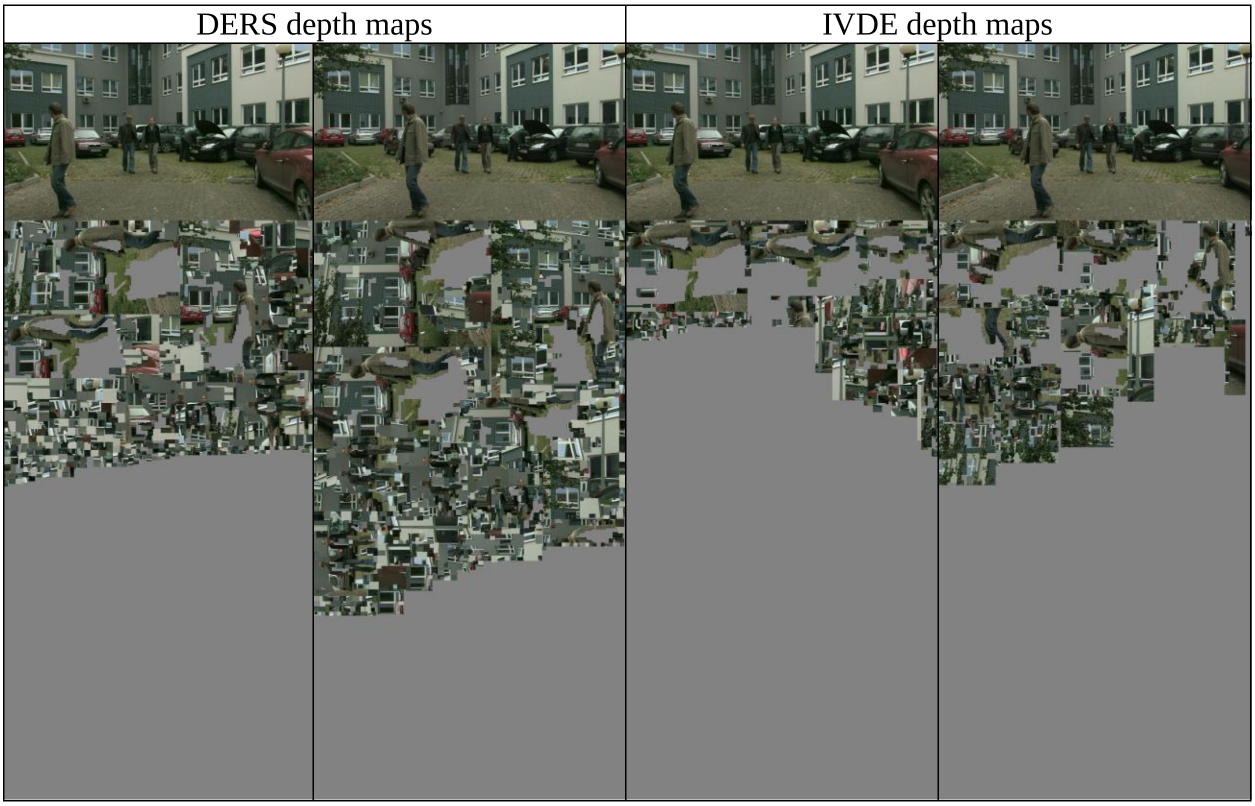


Fig 4. Atlases of TMIV for SP.

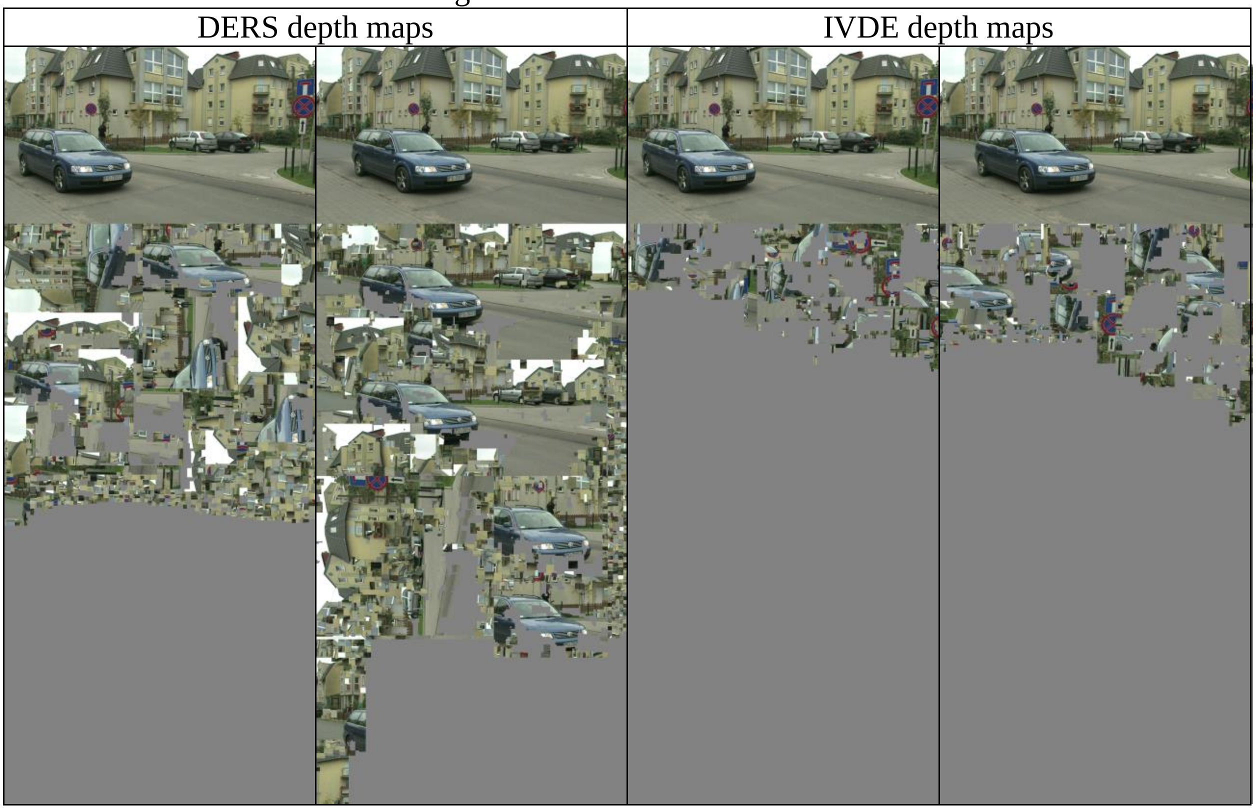


Fig 5. Atlases of TMIV for SU.

3 Conclusions

This document presented a thorough comparison of DERS and IVDE depth estimation methods that tested many aspects of their possible applications.

First of all, the experiment that followed EE Depth CTC have shown slight improvement of quality of IVDE over DERS in most of tested metrics. Moreover, the total time of estimating depth maps was much shorter for IVDE.

In comparison that followed MIV CTC, the IVDE showed much better performance of IVDE, expressed both in objective quality of synthesized input views and in subjective quality of synthesized posetraces. The inter-view consistency of generated depth maps provides much better performance of the TMIV pruner and higher stability of view quality in posetraces. Note that IVDE estimates depth maps also for omnidirectional content. The results of using the IVDE omnidirectional depth maps in SA, SB and SC are presented in [M54278].

4 Recommendations

We recommend to:

- unify Znear and Zfar for all views of SD (1.8 and 4.4, also used earlier in EE Depth),
- change Znear for SL and SG, as some parts of the scene are closer and do not fit in the previous range,
- make IVDE the reference software for depth estimation.

5 Acknowledgement

This work was supported by the Ministry of Science and Higher Education.

6 References

[N19143] Description of DERS, ISO/IEC JTC1/SC29/WG11 MPEG2019/N19143, Brussels, January 2020.

[N19213] Test Model 5 for Immersive Video, ISO/IEC JTC1/SC29/WG11, MPEG2019/N19213, Online, April 2020.

[N19224] Manual of Immersive Video Depth Estimation, ISO/IEC JTC1/SC29/WG11 MPEG2020/N19224, Online, April 2020.

[N19221] Exploration Experiments on Processing for Immersive Video, ISO/IEC JTC1/SC29/WG11 MPEG2019/N19221, Online, April 2020.

[N19214] Common Test Conditions for Immersive Video, ISO/IEC JTC1/SC29/WG11 MPEG2019/N19214, Online, April 2020.

[M53407] D. Mieloch, A. Dziembowski, J. Stankowski, O. Stankiewicz, M. Domański, G. Lee, Y. Jeong, “[MPEG-I Visual] Immersive video depth estimation”, JTC1/SC29/WG11 MPEG2020/M53407, Online, April 2020.

[M53527] J. Sancho, T. Senoh, S. Rogge, D. Bonatto, R. Salvador, E. Juarez, A. Munteanu, G. Lafruit, “[MPEG-I Visual] New configuration parameters in DERS 9.0”, JTC1/SC29/WG11 MPEG2020/M53527, Online, April 2020.

[M54278] D. Mieloch, “IVDE and TMIV CTC depth maps comparison”, JTC1/SC29/WG11 MPEG2020/ M54278, Online, June 2020.