INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 4
MPEG VIDEO CODING

ISO/IEC JTC 1/SC 29/WG 4 m59519
April 2021, Online

Title: **MIV EE-5: Decoder-side depth estimation**

Source: **Adrian Dziembowski (Poznań University of Technology)**

Abstract

The document presents the results of MIV Exploration Experiments 5 on decoder-side depth estimation.

1 Introduction

Owner: Adrian Dziembowski (PUT)

Participants: Jun Young Jeong (ETRI-IM), Adrian Dziembowski (PUT)

EE-5.5: Study of variants of the geometry assistance features

- **Goal**: Test, whether it is more beneficial to send more detailed geometry assistance features for a subset of views, or more generous features for all transmitted views.
- **Description**: Three approaches will be compared, in both the total SEI bitrate should be below 1Mbps:
  - geometry assistance SEI sent for all transmitted views, initial grid size in feature extractor set to 128x128,
  - geometry assistance SEI sent only for views from the first atlas, initial grid size set to 32x32,
  - geometry assistance SEI sent only for views from the first atlas, recursive splitting, initial grid 128x128, max 4 splits.

EE-5.6: Study of input depth assistance in DSDE

- **Goal**: Test, if the DSDE approach with sending of depth maps for a subset of transmitted views can be as effective as the A17 in terms of BD-rates and decoding time.
- **Description**: Two experiments will be conducted:
  - 3 texture atlases, depth sent only for views from first atlas,
  - 3 texture atlases, depth sent only for views from first 2 atlases.

2 Results

2.1 EE-5.5: Study of variants of the geometry assistance features

The experiment was not completed.

2.2 EE-5.6: Study of input depth assistance in DSDE

~~Note: the results below were not crosschecked, as PUT and ETRI-IM used different QP values (PUT used QPs for G17 anchor instead of proper V17 anchor).~~

Update: The experiment was **successfully crosschecked** by ETRI-IM:

- EE-5.6.1: partial crosscheck, sequences D and R,
- EE-5.6.2: partial crosscheck, sequences D, E, and O.

G17 vs. EE-5.6.1 (one geometry atlas):

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	Atlas encoding	Video encoding	Decoding & Rendering	MIV DSDE	#######	Difference [%]	MIV DSDE	#######	Difference [%]
ClassroomVideo	A	-55.9%	-27.3%	-36.8%	-18.2%	1451.9%	121.3%	43.0%	5.69	5.56	-2.4%	4.05	4.14	2.2%
Museum	B	63.5%	29.8%	9.8%	4.7%	1505.5%	153.5%	64.2%	9.18	14.17	54.4%	6.34	10.54	66.3%
Fan	O	54.2%	90.1%	22.3%	62.4%	457.5%	93.3%	57.1%	10.89	9.36	-14.0%	10.03	7.76	-22.6%
Kitchen	J	-5.6%	-10.9%	-14.4%	-13.6%	796.7%	120.2%	53.1%	11.99	12.69	5.8%	11.21	11.53	2.9%
Painter	D	2.4%	20.9%	-9.5%	10.6%	336.3%	134.3%	48.9%	7.60	6.65	-12.5%	7.35	4.37	-40.6%
Frog	E	16.4%	23.1%	22.3%	26.0%	267.3%	105.7%	40.4%	7.40	6.99	-5.5%	7.17	7.16	-0.1%
Carpark	P	51.6%	55.9%	14.6%	29.0%	270.2%	110.7%	53.8%	10.24	9.65	-5.7%	8.19	7.29	-11.1%
Chess	N	---	---	---	---	424.6%	146.6%	70.9%	25.19	31.62	25.5%	23.89	30.25	26.6%
Group	R	---	---	---	---	524.0%	101.9%	50.5%	22.60	24.31	7.6%	23.55	25.18	6.9%
MIV		---	---	---	---	**670.4%**	**120.8%**	**53.5%**	**12.31**	**13.45**	**5.9%**	**11.31**	**12.02**	**3.4%**
Optional content - Proposal vs. Low/High-bitrate Anchors														
Fencing	L	-42.2%	-5.6%	-9.3%	17.8%	277.4%	73.9%	55.7%	12.90	12.58	-2.4%	9.18	8.61	-6.2%
Hall	T	8.9%	-12.7%	-42.4%	-15.3%	282.8%	67.5%	48.7%	16.13	15.59	-3.3%	13.57	12.41	-8.5%
Street	U	6.5%	13.0%	13.9%	20.5%	260.5%	127.3%	55.2%	7.07	6.98	-1.3%	4.91	4.60	-6.3%
ChessPieces	Q	---	---	---	---	433.7%	163.4%	70.0%	27.71	31.86	15.0%	25.79	31.14	20.8%
Hijack	C	---	---	---	---	790.5%	133.7%	62.8%	22.33	25.39	13.7%	21.03	23.01	9.4%
Mirror	I	35.1%	27.5%	20.9%	14.8%	232.1%	134.3%	60.5%	12.41	14.23	14.7%	11.17	12.51	12.0%
Cadillac	G	-8.1%	10.4%	-7.4%	8.4%	484.0%	150.1%	56.8%	14.30	13.95	-2.4%	14.29	14.23	-0.4%
MIV		---	---	---	---	**394.4%**	**121.5%**	**58.5%**	**16.12**	**17.23**	**4.8%**	**14.28**	**15.22**	**3.0%**

G17 vs. EE-5.6.2 (two geometry atlases):

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	Atlas encoding	Video encoding	Decoding & Rendering	MIV DSDE	#######	Difference [%]	MIV DSDE	#######	Difference [%]
ClassroomVideo	A	-28.9%	3.2%	-27.4%	0.9%	1451.5%	109.3%	34.2%	5.69	6.71	17.8%	4.05	4.68	15.5%
Museum	B	-37.7%	-19.9%	-15.9%	-6.5%	1520.2%	152.4%	56.5%	9.18	13.24	44.3%	6.34	9.62	51.7%
Fan	O	96.6%	172.7%	62.5%	128.1%	364.5%	100.2%	38.9%	10.89	9.21	-15.4%	10.03	7.24	-27.8%
Kitchen	J	-14.5%	-8.7%	-18.2%	-10.3%	782.2%	118.6%	45.3%	11.99	12.38	3.2%	11.21	10.81	-3.5%
Painter	D	28.9%	58.0%	14.3%	44.9%	354.6%	157.7%	43.5%	7.60	6.47	-14.9%	7.35	3.67	-50.1%
Frog	E	33.7%	46.0%	38.0%	48.3%	210.4%	117.1%	20.0%	7.40	6.99	-5.5%	7.17	7.19	0.3%
Carpark	P	88.7%	95.0%	46.6%	65.9%	210.2%	114.0%	24.7%	10.24	9.66	-5.6%	8.19	7.54	-7.9%
Chess	N	---	---	---	---	406.6%	165.6%	76.9%	25.19	30.09	19.4%	23.89	29.96	25.4%
Group	R	---	---	---	---	522.0%	107.8%	47.4%	22.60	22.28	-1.4%	23.55	22.85	-3.0%
MIV		---	---	---	---	**646.9%**	**127.0%**	**43.0%**	**12.31**	**13.00**	**4.7%**	**11.31**	**11.51**	**0.1%**
Optional content - Proposal vs. Low/High-bitrate Anchors														
Fencing	L	-23.5%	24.2%	11.0%	47.2%	280.9%	97.2%	15.7%	12.90	12.60	-2.3%	9.18	8.66	-5.7%
Hall	T	29.4%	0.5%	-25.5%	9.8%	349.4%	90.6%	29.1%	16.13	15.36	-4.8%	13.57	11.95	-11.9%
Street	U	21.5%	30.6%	28.5%	37.9%	198.2%	116.3%	45.4%	7.07	7.04	-0.4%	4.91	4.70	-4.1%
ChessPieces	Q	---	---	---	---	454.0%	175.0%	65.0%	27.71	30.29	9.3%	25.79	29.69	15.1%
Hijack	C	---	---	---	---	653.4%	126.5%	49.6%	22.33	25.97	16.3%	21.03	24.44	16.2%
Mirror	I	41.4%	43.8%	22.6%	30.0%	214.6%	142.4%	45.8%	12.41	13.91	12.2%	11.17	12.04	7.8%
Cadillac	G	10.2%	39.7%	8.3%	35.6%	475.7%	162.3%	31.4%	14.30	13.92	-2.6%	14.29	14.24	-0.3%
MIV		---	---	---	---	**375.2%**	**130.0%**	**40.3%**	**16.12**	**17.01**	**4.0%**	**14.28**	**15.10**	**2.4%**

EE-5.6.1 vs. EE-5.6.2:

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	Atlas encoding	Video encoding	Decoding & Rendering	MIV DSDE	#######	Difference [%]	MIV DSDE	#######	Difference [%]
ClassroomVideo	A	68.4%	45.5%	18.1%	24.6%	83.1%	91.5%	69.8%	5.23	6.26	19.7%	3.72	4.10	10.2%
Museum	B	-51.8%	-33.2%	-19.8%	-8.8%	87.6%	86.4%	90.7%	10.16	9.22	-9.2%	7.47	6.61	-11.5%
Fan	O	27.7%	40.9%	24.6%	38.8%	84.0%	129.3%	57.1%	10.23	10.06	-1.6%	8.56	7.98	-6.8%
Kitchen	J	-3.7%	4.0%	-1.9%	4.5%	87.9%	98.4%	89.5%	10.68	10.26	-3.9%	10.04	9.20	-8.3%
Painter	D	22.3%	28.1%	20.9%	27.0%	80.7%	98.4%	50.0%	6.52	6.38	-2.2%	4.20	3.56	-15.3%
Frog	E	13.6%	17.2%	11.6%	16.1%	79.9%	131.4%	36.0%	7.33	7.31	-0.2%	7.48	7.45	-0.4%
Carpark	P	22.7%	25.6%	27.7%	27.9%	79.9%	130.8%	69.8%	9.87	10.24	3.7%	7.52	7.96	5.9%
Chess	N	---	---	---	---	83.1%	94.4%	91.5%	29.30	27.37	-6.6%	27.83	27.11	-2.6%
Group	R	---	-81.8%	---	-70.3%	88.6%	99.9%	67.8%	21.46	19.39	-9.6%	23.56	21.23	-9.9%
MIV		---	---	---	---	**83.9%**	**106.7%**	**69.1%**	**12.31**	**11.83**	**-1.1%**	**11.15**	**10.58**	**-4.3%**
Optional content - Proposal vs. Low/High-bitrate Anchors														
Fencing	L	41.5%	32.5%	18.9%	24.3%	76.1%	96.2%	20.8%	12.88	12.90	0.1%	8.96	9.02	0.6%
Hall	T	8.8%	22.8%	15.6%	48.0%	81.4%	150.5%	40.0%	16.27	16.14	-0.8%	13.54	13.13	-3.0%
Street	U	15.4%	16.4%	12.6%	14.1%	88.1%	117.5%	67.4%	6.98	7.04	0.8%	4.60	4.70	2.3%
ChessPieces	Q	---	---	---	---	78.5%	117.7%	81.6%	30.96	29.17	-5.8%	29.77	28.42	-4.5%
Hijack	C	---	---	---	---	86.7%	112.4%	87.7%	24.10	24.86	3.1%	22.27	23.71	6.5%
Mirror	I	5.8%	15.7%	3.6%	15.2%	86.3%	110.5%	82.0%	12.85	12.61	-1.9%	11.36	11.06	-2.7%
Cadillac	G	13.9%	23.9%	11.1%	23.0%	84.8%	107.2%	56.4%	13.23	13.22	-0.1%	13.48	13.51	0.3%
MIV		---	---	---	---	**83.1%**	**116.0%**	**62.3%**	**16.75**	**16.56**	**-0.6%**	**14.85**	**14.79**	**-0.1%**

Conclusion:

- sending of second, additional geometry atlas is efficient only if cameras are spanned over large angle thus various views are significantly different from each other.

2.3 EE-5.6-related

In this experiment, the equation for deriving QP_D from QP_{TEX} was changed from:

$$QP_D = \max(1, [-14.2 + 0.8 \cdot QP_{TEX}])$$

to:

$$QP_D = \max(1, [0.8 \cdot QP_{TEX}]) .$$

G17 vs. EE-5.6.1 (1 geometry atlas) with changed depth QP:

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	Atlas encoding	Video encoding	Decoding & Rendering	MIV DSDE	#######	Difference [%]	MIV DSDE	#######	Difference [%]
ClassroomVideo	A	-54.4%	-30.9%	-43.2%	-29.0%	1468.3%	88.6%	44.3%	5.69	5.39	-5.4%	4.05	3.94	-2.7%
Museum	B	55.3%	18.1%	3.8%	-4.0%	1901.6%	122.1%	65.6%	9.18	11.40	24.2%	6.34	8.32	31.4%
Fan	O	37.1%	41.4%	-1.1%	13.9%	367.9%	114.0%	63.5%	10.89	10.04	-7.8%	10.03	8.35	-16.7%
Kitchen	J	4.0%	-1.7%	-14.3%	-15.2%	804.2%	101.5%	59.1%	11.99	11.59	-3.3%	11.21	10.63	-5.1%
Painter	D	-24.9%	-16.3%	-37.4%	-26.0%	358.6%	113.9%	59.4%	7.60	6.51	-14.4%	7.35	4.13	-43.8%
Frog	E	7.3%	7.7%	11.5%	9.8%	285.4%	101.1%	43.2%	7.40	7.37	-0.4%	7.17	7.55	5.4%
Carpark	P	26.4%	25.0%	-5.5%	0.6%	214.4%	115.2%	47.2%	10.24	9.85	-3.8%	8.19	7.55	-7.8%
Chess	N	---	---	---	---	412.1%	106.8%	75.9%	25.19	29.37	16.6%	23.89	28.09	17.6%
Group	R	---	---	---	---	558.3%	85.6%	58.9%	22.60	21.48	-4.9%	23.55	23.56	0.0%
MIV		---	---	---	---	**707.8%**	**105.4%**	**57.5%**	**12.31**	**12.56**	**0.1%**	**11.31**	**11.35**	**-2.4%**
Optional content - Proposal vs. Low/High-bitrate Anchors														
Fencing	L	-72.3%	-30.8%	-24.8%	-9.0%	360.5%	94.9%	52.1%	12.90	12.87	-0.2%	9.18	8.94	-2.6%
Hall	T	-18.9%	-66.0%	---	-82.4%	278.5%	103.9%	61.7%	16.13	16.07	-0.4%	13.57	13.17	-3.0%
Street	U	-13.8%	-12.8%	-6.5%	-7.8%	439.4%	130.7%	57.7%	7.07	7.01	-0.9%	4.91	4.58	-6.5%
ChessPieces	Q	---	---	---	---	435.1%	115.4%	70.5%	27.71	30.74	10.9%	25.79	29.91	16.0%
Hijack	C	---	---	---	---	785.7%	96.1%	56.6%	22.33	24.13	8.0%	21.03	22.32	6.1%
Mirror	I	17.3%	8.4%	7.2%	-1.7%	228.1%	113.9%	66.0%	12.41	12.86	3.7%	11.17	11.40	2.0%
Cadillac	G	-14.6%	-4.4%	-19.8%	-11.7%	479.6%	137.8%	52.4%	14.30	13.19	-7.7%	14.29	13.45	-5.9%
MIV		---	---	---	---	**429.6%**	**113.3%**	**59.6%**	**16.12**	**16.69**	**1.9%**	**14.28**	**14.82**	**0.9%**

EE-5.6.1 vs. EE-5.6.1 with changed depth QP:

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	Atlas encoding	Video encoding	Decoding & Rendering	MIV DSDE	#######	Difference [%]	MIV DSDE	#######	Difference [%]
ClassroomVideo	A	3.7%	-4.7%	-7.7%	-10.9%	84.6%	98.3%	97.1%	5.23	5.39	3.0%	3.72	3.94	5.8%
Museum	B	1.4%	-4.9%	-5.7%	-8.8%	108.3%	108.7%	97.6%	10.16	11.40	12.2%	7.47	8.32	11.5%
Fan	O	-11.9%	-23.6%	-16.7%	-26.3%	82.9%	119.1%	95.7%	10.23	10.04	-1.8%	8.56	8.35	-2.5%
Kitchen	J	13.2%	9.4%	-1.2%	-2.8%	90.2%	126.5%	116.1%	10.68	11.59	8.6%	10.04	10.63	5.9%
Painter	D	-22.4%	-27.4%	-22.9%	-27.7%	104.6%	88.2%	93.8%	6.52	6.51	-0.2%	4.20	4.13	-1.7%
Frog	E	-8.0%	-11.5%	-9.1%	-12.2%	102.2%	108.9%	89.8%	7.33	7.37	0.7%	7.48	7.55	0.9%
Carpark	P	-17.5%	-18.6%	-17.0%	-19.1%	79.3%	122.6%	96.3%	9.87	9.85	-0.3%	7.52	7.55	0.4%
Chess	N	#######	492.5%	51.0%	93.3%	83.8%	100.0%	95.7%	29.30	29.37	0.2%	27.83	28.09	0.9%
Group	R	4.2%	0.5%	-0.1%	-4.9%	88.9%	105.3%	103.0%	21.46	21.48	0.1%	23.56	23.56	0.0%
MIV		**######**	**45.8%**	**-3.3%**	**-2.1%**	**91.6%**	**108.6%**	**98.3%**	**12.31**	**12.56**	**2.5%**	**11.15**	**11.35**	**2.4%**
Optional content - Proposal vs. Low/High-bitrate Anchors														
Fencing	L	-19.7%	-19.2%	-14.8%	-19.0%	100.4%	83.8%	83.4%	12.88	12.87	-0.1%	8.96	8.94	-0.3%
Hall	T	-23.5%	-39.9%	-32.6%	-44.5%	74.4%	75.0%	110.3%	16.27	16.07	-1.3%	13.54	13.17	-2.8%
Street	U	-16.4%	-19.5%	-16.5%	-20.9%	165.3%	93.8%	84.5%	6.98	7.01	0.4%	4.60	4.58	-0.3%
ChessPieces	Q	49.5%	4639.1%	16.2%	-11.6%	78.8%	104.1%	92.2%	30.96	30.74	-0.7%	29.77	29.91	0.5%
Hijack	C	91.3%	37.3%	40.9%	12.9%	104.2%	103.2%	103.3%	24.10	24.13	0.1%	22.27	22.32	0.2%
Mirror	I	-11.0%	-14.8%	-10.6%	-15.8%	81.7%	100.4%	99.9%	12.85	12.86	0.1%	11.36	11.40	0.3%
Cadillac	G	-3.2%	-11.6%	-9.6%	-16.6%	85.3%	95.1%	94.7%	13.23	13.19	-0.3%	13.48	13.45	-0.2%
MIV		**9.6%**	**653.1%**	**-3.9%**	**-16.5%**	**98.6%**	**93.6%**	**95.5%**	**16.75**	**16.69**	**-0.3%**	**14.85**	**14.82**	**-0.4%**

3 Recommendations:

PUT recommends:

- to stop EE-5.5 and EE-5.6,
- to use modified QP_D when using input depth assistance in DSDE.

Acknowledgement

The research was supported by the Ministry of Science and Higher Education of Republic of Poland.