

**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 m 55748
October 2020, Online

Title: Exploration Experiments on Future MIV: PUT results
Source: Dawid Mieloch, Adrian Dziembowski (Poznań University of Technology)

Abstract & Recommendations

The documents presents results of EE-related experiments that were conducted by PUT. The results include EE1a, EE2a, EE2b, EE4 and EE5. The summarized recommendations are:

- EE2.a should be continued to test new possible configurations of the feature extractor.
- EE2.b should not be continued, as provided results show that the use of all features provides (on average) better quality of decoder-derived depth maps with additional complexity reduction.
- EE4: we recommend to re-discuss the possibility of substituting HM with x265 in MIV-related experiments.
- EE5 should be continued to test the performance of the new TMIV 8.0 and new (possible) IVDE 3.0.

1 Introduction

The documents presents results of EE-related experiments that were conducted by PUT. The results include partial results for EE1a and full results for EE2a, EE2b, EE4 and EE5. All results can be found in attached reporting templates.

2 Experiments

For all experiments GCC 10.2.0 was used to compile all used software.

2.1 EE1.a

This experiment tested the use of VVC instead of HEVC for the encoding of atlases. The table below compares A17 anchor against the results acquired in this experiment.

Mandatory content - Proposal vs. Low/High-bitrate Anchors								Runtime ratio (%)							
Sequence		High-BR	Low-BR	Max delta	High-BR	Low-BR	High-BR	Low-BR	Pixel rate [%]	Pixel rate [GP/s]	Frame rate [Hz]	TMIV	HM	HM	TMIV
		BD rate Y-PSNR	BD rate Y-PSNR		BD rate VMAF	BD rate VMAF	BD rate IV-PSNR	BD rate IV-PSNR				encoding	encoding	decoding	decoding
Kitchen	SJ	-19.7%	-25.3%	16.37	-27.9%	-30.9%	-24.6%	-23.0%	62%	0.67	30	374.4%	434.2%	#####	102.3%
Carpark	SP	-29.4%	-28.8%	7.51	-30.6%	-31.7%	-16.4%	-19.6%	52%	0.56	25	502.3%	513.3%	#####	100.2%
Group	SR	-36.5%	-34.0%	12.19	-39.2%	-37.1%	-27.3%	-25.1%	62%	0.67	30	471.7%	531.0%	#####	100.4%
MIV		-28.5%	-29.4%	12.02	-32.5%	-33.2%	-22.8%	-22.6%	59%	0.63		449.5%	492.8%	#####	100.9%

PUT was calculating EE1.a for SP, SJ and SR. The exact match was acquired for cross-check of the SJ and SR.

2.2 EE2

2.2.1 EE2.a:

This experiment shows the performance of the decoder-side depth estimation when the encoder-derived auxiliary information is utilized. In this experiment, the modified IVDE used skip and partition flags, and z_{\min} , z_{\max} values derived in the encoder from the CTC depth maps. The table below compares the G17 anchor with this experiment.

Mandatory content - Proposal vs. Low/High-bitrate Anchors								Runtime ratio (%)							
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 [%]	Pixel rate [GP/s]	Frame rate [Hz]	TMIV encoding	HM encoding	HM decoding	TMIV decoding
ClassroomVideo	SA	---	-76.8%	5.66	---	---	-21.1%	-11.3%	106%	1.14	30	100.0%	100.0%	#####	9.7%
Museum	SB	---	---	13.29	-48.1%	-29.3%	-59.1%	-33.3%	106%	1.14	30	100.0%	100.0%	#####	8.2%
Fan	SO	-0.4%	3.2%	10.13	4.0%	7.1%	-0.0%	4.1%	106%	1.14	30	100.0%	100.0%	#####	14.6%
Kitchen	SJ	-28.4%	-20.4%	12.82	-15.3%	-9.2%	-24.5%	-17.2%	106%	1.14	30	100.0%	100.0%	#####	18.2%
Painter	SD	-8.9%	21.3%	8.18	72.5%	71.4%	-21.3%	10.3%	106%	1.14	30	100.0%	100.0%	#####	20.7%
Frog	SE	3.5%	13.7%	6.58	10.0%	19.2%	-1.0%	12.6%	106%	1.14	30	100.0%	100.0%	#####	13.8%
Carpark	SP	-24.0%	-10.5%	9.59	-15.5%	-3.9%	13.1%	12.7%	89%	0.95	25	100.0%	100.0%	#####	20.7%
Chess	SN	---	---	23.54	-67.2%	-43.1%	---	---	106%	1.14	30	100.0%	100.0%	#####	6.7%
Group	SR	---	---	17.93	-62.2%	-39.4%	---	---	106%	1.14	30	100.0%	100.0%	#####	17.4%
MIV		---	---	11.97	---	---	---	---	104%	1.12		100.0%	100.0%	#####	14.5%

Optional content - Proposal vs. Low/High-bitrate Anchors								Runtime ratio (%)							
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 [%]	Pixel rate [GP/s]	Frame rate [Hz]	TMIV encoding	HM encoding	HM decoding	TMIV decoding
Fencing	SL	-35.8%	-15.1%	12.68	-3.0%	6.1%	-3.6%	1.2%	89%	0.95	25	100.0%	100.0%	#####	14.4%
Hall	SU	14.2%	14.3%	8.12	4.0%	8.9%	8.0%	10.3%	89%	0.95	25	100.0%	100.0%	#####	20.8%
Street	ST	---	---	17.25	133.0%	59.4%	---	---	89%	0.95	25	100.0%	100.0%	#####	12.8%
ChessPieces	SQ	---	---	30.94	118.7%	43.3%	---	---	106%	1.14	30	100.0%	100.0%	#####	8.7%
Hijack	SC	---	---	20.96	---	-57.7%	---	---	106%	1.14	30	100.0%	100.0%	#####	13.6%
MIV		---	---	17.99	---	12.0%	---	---	96%	1.02		100.0%	100.0%	#####	14.1%

In cross-check, the TMIV 7.0.1 was compared to 7.0RC5, so there were very big differences in SN and SQ (7.0.1 fixed a bug in G17 that was in these sequences in 7.0). Other sequences are exactly the same, besides some small differences in SR in synthesized views (SA showed -0.3% in PSNR but bitrates and quality are the same).

Comments:

- The encoder-derived features provided a decrease of the BD-rate for almost all mandatory sequences, together with a significant reduction of time required by the IVDE to estimate depth maps.
- The performance of the feature-driven IVDE is dependent on the quality of depth maps that were used to derive the features. The decrease of BD-rate is on average much higher for CGI sequences than for natural content.
- The quality increase is larger for high bitrates.

Recommendations:

- EE2.a should be continued to test new possible configurations of the feature extractor.

2.2.2 EE2.b:

This experiment shows the performance of the decoder-side depth estimation when only the encoder-derived skip flag is used. The table below compares the G17 anchor with this experiment.

Mandatory content - Proposal vs. Low/High-bitrate Anchors								Runtime ratio (%)							
Sequence		High-BR	Low-BR	Max delta	High-BR	Low-BR	High-BR	Low-BR	Pixel rate [%]	Pixel rate [GP/s]	Frame rate [Hz]	TMIV	HM	HM	TMIV
		BD rate Y-PSNR	BD rate Y-PSNR		BD rate VMAF	BD rate VMAF	BD rate IV-PSNR	BD rate IV-PSNR				encoding	encoding	decoding	decoding
ClassroomVideo	SA	92.9%	27.1%	5.68	157.6%	24.1%	9.4%	8.3%	106%	1.14	30	100.0%	100.0%	#####	11.9%
Museum	SB	-99.9%	-0.5%	14.84	-8.4%	-1.1%	-20.0%	-2.0%	106%	1.14	30	100.0%	100.0%	#####	10.1%
Fan	SO	5.2%	3.8%	10.31	3.2%	2.4%	5.3%	4.3%	106%	1.14	30	100.0%	100.0%	#####	26.8%
Kitchen	SJ	10.0%	8.7%	13.46	8.1%	5.4%	8.7%	8.0%	106%	1.14	30	100.0%	100.0%	#####	24.8%
Painter	SD	-43.8%	-26.6%	7.57	-21.0%	-10.4%	-39.3%	-24.2%	106%	1.14	30	100.0%	100.0%	#####	79.7%
Frog	SE	-2.5%	-1.5%	6.62	-0.8%	-0.4%	-3.8%	-2.0%	106%	1.14	30	100.0%	100.0%	#####	60.9%
Carpark	SP	-5.8%	-4.3%	9.57	-5.6%	-3.3%	4.6%	1.9%	89%	0.95	25	100.0%	100.0%	#####	81.6%
Chess	SN	-5.4%	-5.7%	25.13	-9.2%	-5.1%	-65.7%	0.9%	106%	1.14	30	100.0%	100.0%	#####	6.8%
Group	SR	---	-89.3%	22.41	-62.3%	-39.3%	---	-99.6%	106%	1.14	30	100.0%	100.0%	#####	36.7%
MIV		---	-9.8%	12.84	6.8%	-3.1%	---	-11.6%	104%	1.12		100.0%	100.0%	#####	37.7%

Optional content - Proposal vs. Low/High-bitrate Anchors								Runtime ratio (%)							
Sequence		High-BR	Low-BR	Max delta	High-BR	Low-BR	High-BR	Low-BR	Pixel rate [%]	Pixel rate [GP/s]	Frame rate [Hz]	TMIV	HM	HM	TMIV
		BD rate Y-PSNR	BD rate Y-PSNR		BD rate VMAF	BD rate VMAF	BD rate IV-PSNR	BD rate IV-PSNR				encoding	encoding	decoding	decoding
Fencing	SL	17.9%	18.9%	12.75	6.4%	3.9%	10.0%	9.1%	89%	0.95	25	100.0%	100.0%	#####	37.8%
Hall	SU	-0.5%	-0.6%	7.99	-0.6%	-0.5%	0.6%	0.3%	89%	0.95	25	100.0%	100.0%	#####	84.8%
Street	ST	---	---	18.00	---	82.7%	---	---	89%	0.95	25	100.0%	100.0%	#####	64.9%
ChessPieces	SQ	-41.6%	-54.7%	29.64	5.0%	0.2%	-13.1%	-22.0%	106%	1.14	30	100.0%	100.0%	#####	9.0%
Hijack	SC	48.3%	26.6%	22.82	17.7%	7.7%	31.2%	14.0%	106%	1.14	30	100.0%	100.0%	#####	17.1%
MIV		---	---	18.24	---	18.8%	---	---	96%	1.02		100.0%	100.0%	#####	42.7%

In cross-check, the 7.0.1 TMIV was compared with 7.0RC5, so differences in SN and SQ occurred. Other sequences were exactly the same.

Comments:

- The use of skip flag only showed a smaller decrease of the computational time than in the previous experiment (about 3 times slower depth estimation), as the use of encoder-derived z_{min} and z_{max} significantly decreases the complexity of the estimation.
- In mandatory natural sequences (SD, SE, and SP) the skip-flag only mode showed better performance than the use of all features. If the features are derived from estimated (non-rendered) depth maps, then errors in input depth maps seem to negatively impact the quality of decoder-derived depth maps.

Recommendations:

- EE2.b should not be continued, as provided results show that the use of all features provides (on average) better quality of decoder-derived depth maps with additional complexity reduction.

2.3 EE4

The experiment compares performance of MIV using two video encoder implementations:

- HM 16.16 (as in MIV CTC),
- x265 (libx265 in ffmpeg-4.3.1).

Mandatory content - Proposal vs. Low/High-bitrate Anchors								Runtime ratio (%)							
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 [%]	Pixel rate [GP/s]	Frame rate [Hz]	TMIV encoding	HM encoding	HM decoding	TMIV decoding
ClassroomVideo	SA	67.3%	61.1%	3.26	-6.5%	17.5%	21.8%	28.1%	63%	0.67	30	18.8%	0.2%	89.9%	79.0%
Museum	SB	-38.2%	-31.9%	15.73	-35.6%	-31.4%	-40.0%	-37.8%	63%	0.67	30	25.4%	0.1%	100.0%	86.1%
Fan	SO	64.5%	71.1%	7.26	54.3%	61.3%	59.5%	63.7%	62%	0.67	30	7.2%	0.2%	100.0%	66.0%
Kitchen	SJ	1.6%	9.1%	16.13	2.3%	11.2%	3.1%	7.4%	62%	0.67	30	14.3%	0.2%	100.0%	71.7%
Painter	SD	55.6%	59.5%	8.19	40.9%	46.9%	41.3%	50.0%	63%	0.67	30	6.7%	0.2%	100.0%	72.6%
Frog	SE	52.2%	54.8%	6.57	32.0%	38.7%	37.1%	42.5%	62%	0.67	30	3.8%	0.2%	100.0%	66.9%
Carpark	SP	54.6%	57.2%	8.31	33.9%	43.9%	52.4%	48.4%	52%	0.56	25	3.0%	0.2%	100.0%	66.2%
Chess	SN	13.1%	23.5%	15.74	9.8%	22.7%	21.5%	27.9%	63%	0.67	30	11.0%	0.1%	100.0%	74.1%
Group	SR	3.7%	23.0%	11.68	-1.2%	17.5%	2.1%	15.2%	62%	0.67	30	10.3%	0.2%	100.0%	67.0%
MIV		30.5%	36.4%	10.32	14.4%	25.4%	22.1%	27.3%	61%	0.66		11.2%	0.2%	98.9%	72.2%

Optional content - Proposal vs. Low/High-bitrate Anchors								Runtime ratio (%)							
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 [%]	Pixel rate [GP/s]	Frame rate [Hz]	TMIV encoding	HM encoding	HM decoding	TMIV decoding
Fencing	SL	66.9%	62.4%	11.72	36.7%	46.5%	45.0%	48.4%	52%	0.56	25	3.0%	0.2%	100.0%	68.8%
Street	SU	40.9%	36.3%	10.72	2.6%	17.5%	41.8%	31.9%	52%	0.56	25	3.0%	0.2%	100.0%	67.4%
Hall	ST	69.8%	64.3%	11.40	43.7%	48.7%	60.5%	70.8%	52%	0.56	25	2.4%	0.2%	100.0%	66.4%
ChessPieces	SQ	3.6%	20.1%	15.97	7.9%	22.5%	35.2%	38.8%	63%	0.67	30	9.0%	0.2%	100.0%	67.2%
Hijack	SC	55.2%	59.7%	9.82	43.6%	50.6%	70.8%	68.1%	63%	0.67	30	12.1%	0.2%	100.0%	79.8%
MIV		47.3%	48.6%	11.93	26.9%	37.2%	50.7%	51.6%	56%	0.60		5.9%	0.2%	100.0%	69.9%

Comments:

- The use of the x265 encoder reduces the video encoding step by a factor of 500, which allows reducing the time needed for entire TMIV encoding 10 times.
- The encoding efficiency is lower when using x265 for almost all the content (except for SB, where it decreases the bitrate without any quality decrease).

Recommendations:

- We recommend to re-discuss the possibility of substituting HM with x265 in MIV-related experiments.

2.4 EE5

The experiment tested the performance of depth maps estimated by IVDE in comparison with the current CTC depth maps. The table below compares the performance of the A17 anchor against the new depth maps (estimated at the TMIV encoder side).

Mandatory content - Proposal vs. Low/High-bitrate Anchors								Runtime ratio (%)							
Sequence		High-BR	Low-BR	Max	High-BR	Low-BR	High-BR	Low-BR	Pixel	Pixel	Frame	TMIV	HM	HM	TMIV
		BD rate	BD rate		BD rate	BD rate	BD rate	BD rate							
		Y-PSNR	Y-PSNR	Y-PSNR	VMAF	VMAF	IV-PSNR	IV-PSNR	[%]	[GP/s]	[Hz]				
ClassroomVideo	SA	---	---	4.14	---	---	688.0%	633.9%	63%	0.67	30	164.5%	177.7%	100.0%	99.2%
Museum	SB	---	---	25.09	---	---	---	---	63%	0.67	30	92.3%	72.2%	100.0%	92.0%
Fan	SO	-68.0%	-67.0%	5.62	-55.5%	-58.6%	-51.7%	-55.3%	62%	0.67	30	112.2%	114.2%	100.0%	82.8%
Kitchen	SJ	154.4%	94.2%	15.79	202.3%	101.6%	88.2%	62.8%	62%	0.67	30	133.6%	140.3%	100.0%	81.1%
Painter	SD	67.2%	63.2%	7.57	63.1%	61.8%	79.8%	68.9%	63%	0.67	30	94.6%	95.0%	100.0%	75.9%
Frog	SE	-4.8%	-1.1%	5.33	-2.8%	-0.4%	0.2%	1.5%	62%	0.67	30	95.8%	96.4%	100.0%	73.3%
Carpark	SP	47.3%	60.2%	6.99	39.9%	57.4%	51.1%	62.1%	52%	0.56	25	81.7%	81.4%	100.0%	61.4%
Chess	SN	---	---	27.26	---	---	---	---	63%	0.67	30	71.9%	66.3%	100.0%	68.5%
Group	SR	---	---	28.92	---	145.7%	---	---	62%	0.67	30	122.3%	67.7%	100.0%	78.9%
MIV		---	---	14.08	---	---	---	---	61%	0.66		107.7%	101.2%	100.0%	79.2%

Optional content - Proposal vs. Low/High-bitrate Anchors								Runtime ratio (%)							
Sequence		High-BR	Low-BR	Max	High-BR	Low-BR	High-BR	Low-BR	Pixel	Pixel	Frame	TMIV	HM	HM	TMIV
		BD rate	BD rate		BD rate	BD rate	BD rate	BD rate							
		Y-PSNR	Y-PSNR	Y-PSNR	VMAF	VMAF	IV-PSNR	IV-PSNR	[%]	[GP/s]	[Hz]				
Fencing	SL	1.0%	26.7%	10.15	26.7%	38.2%	47.6%	48.6%	52%	0.56	25	107.8%	108.3%	100.0%	72.1%
Street	SU	26.7%	37.9%	8.58	11.0%	30.8%	23.9%	36.1%	52%	0.56	25	135.5%	136.9%	100.0%	64.8%
Hall	ST	-54.4%	-47.1%	9.47	-53.3%	-47.5%	-46.5%	-42.8%	52%	0.56	25	111.6%	112.7%	100.0%	65.9%
ChessPieces	SQ	---	---	27.82	---	---	---	---	63%	0.67	30	98.9%	93.3%	100.0%	72.1%
Hijack	SC	---	---	20.66	---	186.2%	---	---	63%	0.67	30	76.3%	73.5%	100.0%	78.3%
MIV		---	---	15.34	---	---	---	---	56%	0.60		106.0%	105.0%	100.0%	70.6%

During cross-check of depth maps some differences were identified. Problem is described in <http://mpegx.int-evry.fr/software/MPEG/Explorations/6DoF/IVDE/-/issues/13>, most of differences will be fixed in IVDE 3.0.

Cross-check of TMIV encoding was performed using the depth maps provided by PUT. We identified differences in anchor computation for different compilers (different atlases computed by TMIV encoder), a short test with VC16 on Windows 10 has given matching results. The compiler-related differences in EPR sequences occur both in TMIV and IVDE, for both of them some work was already done to fix it.

Comments:

- As expected, the quality of depth maps generated in the experiment is lower than for CTC depth maps. The depth maps in this experiment are generated using the same estimation parameters for all sequences, while for CTC depth maps (even if they were generated earlier using IVDE), the parameters were fine-tuned to give the best possible quality.
- The high quality in SO is the result of much higher redundancy in atlases when estimated depth maps are used (more information from input views is transmitted, resulting in the increased quality of synthesized views). There are also fewer high-frequency edges in depth maps (fewer details on a fan), which decreased the bitrate of encoded geometry atlases.
- A high BD-rate decrease was observed for ST. The possibility of generating new CTC depth maps for this sequence will be considered.

Recommendations:

- EE5 should be continued to test the performance of the new TMIV 8.0 and new (possible) IVDE 3.0.

Acknowledgement

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