INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 04
MPEG VIDEO CODING

ISO/IEC JTC 1/SC 29/WG 04 m 58495
January 2022, Online

Title: **Outcomes of exploration experiments – summary report**

Source: **J. Jung (Tencent), D. Mieloch (Poznan University of Technology), S. Fachada (ULB), L. Ciccarelli (V-Nova), A. Dziembowski (Poznan University of Technology)**

Abstract

This contribution is a summary of outcomes of all experiments listed in N0145. A total of 10 organizations participated in one or more of the listed experiments. Six main experiments were agreed upon, with all except EE-1 having additional sub-experiments. Significant participation and engagement from experimenters were observed, and several useful recommendations are provided from participating organizations.

Introduction

Six main exploration experiments, most having additional sub-experiments, were agreed upon in MPEG-136. The summary in this contribution is collated from detailed reports from experimenters produced in documents listed in Table 1.

Table 1: Input document from experimenters

m58479	Tencent results for Exploration Experiments on Coding for Future MPEG Immersive Video
m58564	ETRI-IM results for Exploration Experiments on Future MIV
m58835	Exploration Experiments on Future MIV: PUT results
m58968	Result of experiment using LCEVC in TMIV

EE1: IVDE depth maps generation

Owner: Dawid Mieloch (PUT)

Description: This experiment generates a MIV anchor based on depth maps obtained with IVDE 5.0 with features extracted internally from source textures.

Participants: Jun Young Jeong (ETRI-IM), Dawid Mieloch (PUT), Yupeng Xie (ULB), Eduardo Juarez (UPM)

Cross-check: The MIV part was successfully cross-checked for all sequences (with minor differences below 0.2%). The cross-check of the IVDE part was performed for sequences A, B, C, D, E, G, I, and L and has shown minor differences in E and more noticeable one for C. It is very likely that they come from the different version of the gcc compiler employed. Differences were shown in comment http://mpegx.int-evry.fr/software/MPEG/MIV/InputDocuments/-/issues/413#note_52884. The cross-check for other sequences was not reported by Yupeng Xie.

Results:

The table below shows the comparison of MIV A17 anchor with CTC depth maps and with depth maps estimated in this EE:

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 Anchor	EE1	Difference [%]	MIV Anchor	EE1	Difference [%]
ClassroomVideo	A	974,7%	209,7%	193,8%	146,6%	0%	0,00	30	111,5%	162,6%	109,2%	0,99	2,65	168,9%	0,76	1,23	62,3%
Museum	B	---	---	---	467,7%	0%	0,00	30	165,4%	148,3%	120,4%	9,45	18,75	98,6%	5,35	16,59	209,9%
Fan	O	-75,2%	-70,7%	-50,5%	-47,3%	0%	0,00	30	81,5%	157,9%	142,6%	8,02	6,12	-23,6%	7,24	6,70	-7,4%
Kitchen	J	145,9%	76,1%	126,9%	61,8%	0%	0,00	30	87,6%	120,9%	118,4%	14,67	14,77	0,6%	11,19	11,75	5,0%
Painter	D	1,1%	-0,3%	4,1%	1,3%	0%	0,00	30	128,3%	99,7%	108,7%	7,94	7,50	-5,6%	5,26	5,58	6,1%
Frog	E	-20,6%	-12,6%	-12,1%	-7,9%	0%	0,00	30	109,6%	101,6%	108,3%	7,39	6,36	-13,9%	7,21	5,89	-18,3%
Carpark	P	0,6%	3,7%	3,0%	5,0%	0%	0,00	25	98,5%	72,6%	104,0%	7,05	6,99	-0,9%	5,01	4,96	-1,1%
Chess	N	---	---	---	---	0%	0,00	30	162,1%	93,0%	112,4%	13,60	28,33	108,3%	12,44	27,38	120,1%
Group	R	---	---	---	316,2%	0%	0,00	30	172,6%	77,3%	111,1%	12,89	22,09	71,4%	10,30	20,33	97,4%
MIV		**---**	**---**	**---**	**---**	**0%**	**0,00**		**124,1%**	**114,9%**	**115,0%**	**9,11**	**12,62**	**44,9%**	**7,20**	**11,16**	**52,7%**
Optional content - Proposal vs. Low/High-bitrate Anchors																	
Fencing	L	5,0%	14,0%	-16,5%	7,4%	0%	0,00	25	108,4%	105,2%	108,8%	10,37	9,54	-8,0%	7,60	4,15	-45,4%
Hall	T	-62,3%	-48,5%	-44,8%	-39,8%	0%	0,00	25	100,0%	69,2%	93,1%	11,67	10,05	-13,8%	8,27	7,75	-6,2%
Street	U	-5,3%	-4,8%	-10,4%	-6,4%	0%	0,00	25	116,1%	95,5%	113,9%	8,48	8,52	0,5%	4,54	4,48	-1,4%
ChessPieces	Q	---	---	---	---	0%	0,00	30	123,4%	95,8%	105,6%	14,44	33,74	133,7%	15,29	34,00	122,4%
Hijack	C	---	---	---	---	0%	0,00	30	115,5%	83,4%	105,5%	7,98	21,49	169,2%	5,70	19,97	250,4%
Mirror	I	-6,0%	-13,1%	-6,2%	-13,6%	0%	0,00	30	99,2%	80,4%	104,7%	8,76	9,50	8,5%	5,23	6,10	16,6%
Cadillac	G	-0,3%	-15,0%	17,1%	-0,8%	0%	0,00	30	87,5%	101,7%	117,6%	12,08	12,93	7,0%	11,16	11,27	1,0%
MIV		**---**	**---**	**---**	**---**	**0%**	**0,00**		**107,2%**	**90,2%**	**107,0%**	**10,54**	**15,11**	**42,4%**	**8,26**	**12,53**	**48,2%**

Recommendations:

ETRI:

- Maintain the current CTC depth maps without any replacement.

PUT:

- No change to CTC depth maps due to too small differences in posetraces.
- Continue the EE1.

EE2: verification tests preparation

Owner: Dawid Mieloch (PUT)

Description: With a view of producing anchors for the verification tests, the goal of this experiment was to refine simulation pipeline from the previous meeting cycle and have an initial

performance evaluation of using the Multi-View High Efficiency Video Codec (MV-HEVC). For this experiment, only sequences that never used for the MIV development were evaluated.

Participants: Dawid Mieloch (PUT), Bart Kroon (Philips), Jun Young Jeong (ETRI-IM), Franck Thudor (InterDigital)

Cross-check: The cross-check was successful both for EE2.1 and EE2.2. One of the sequences (Cyberpunk – X) was added after the description of the EEs was finalized, therefore, was not cross-checked.

Results:

3 posetraces for each of 7 sequences and 5 rate points can be found on the content server:

- MV-HEVC + RVS: MPEG-I/Part12-ImmersiveVideo/for_testing/N0145_EE2.1/
- MIV: MPEG-I/Part12-ImmersiveVideo/for_testing/N0145_EE2.2/
- MIV best-reference: MPEG-I/Part12-ImmersiveVideo/for_testing/N0145_EE2_R97/

Sequences are:

- F (Guitarist)
- H (BabyUnicorn)
- K (Breaktime)
- W (Dancing)
- X (Cyberpunk)
- Y (Barn)
- Z (Breakfast)

QP for geometry for MIV is computed with the formula in the MIV CTC. Tuned QPs for textures for MIV (EE2.2) are:

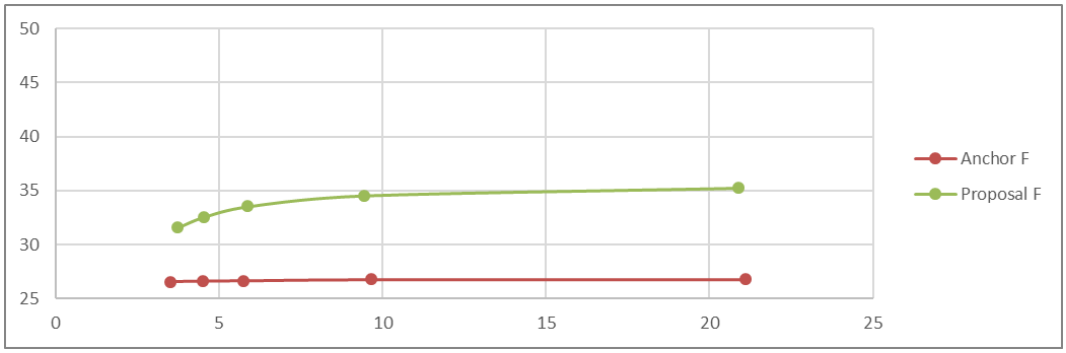
- F [29, 38, 44, 48, 51]
- H [28, 34, 40, 46, 51]
- K [23, 30, 37, 44, 51]
- W [23, 30, 37, 44, 51]
- X [19, 25, 31, 36, 44]
- Y [23, 30, 37, 44, 51]
- Z [23, 30, 37, 44, 51]

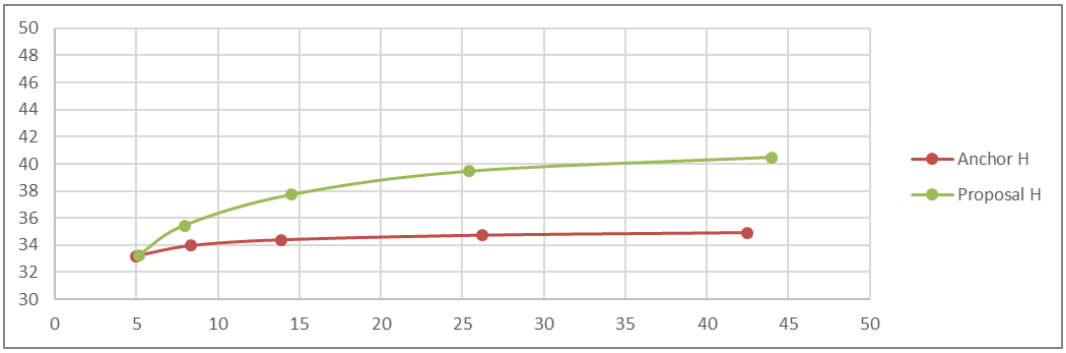
QP for geometry for MV-HEVC is computed with the formula qp[geo]=qp[tex]-10. Tuned QPs for textures for MV-HEVS (EE2.1) are:

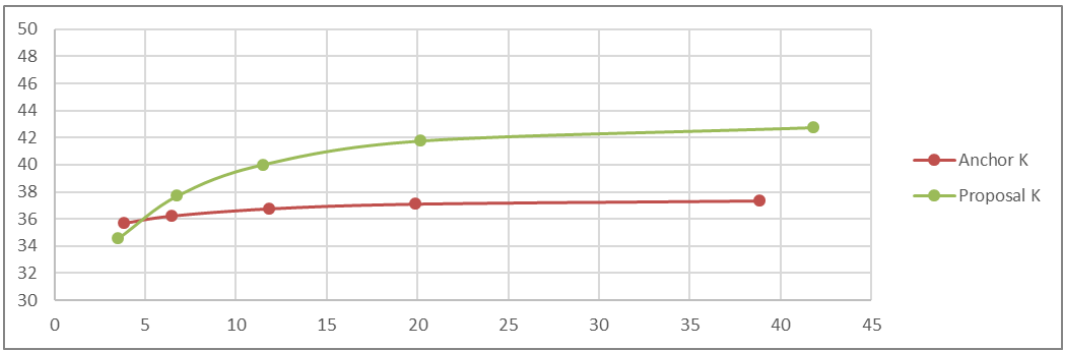
- F [21, 27, 31, 33, 35]
- H [25, 29, 34, 38, 42]

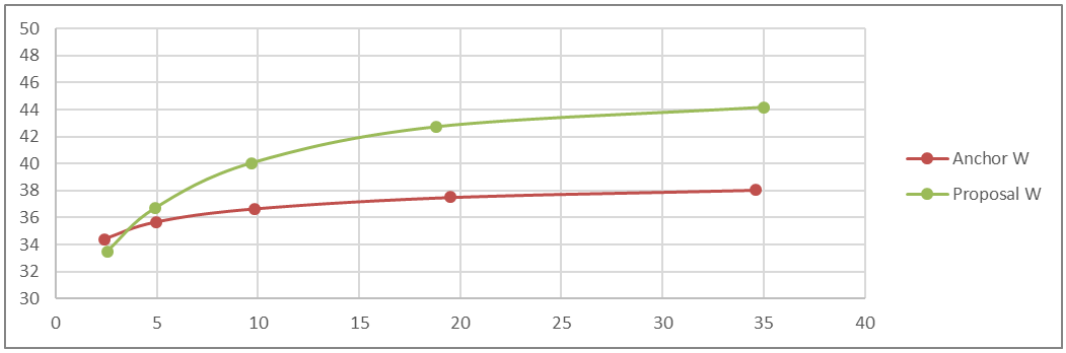
- K [18, 23, 28, 33, 37]
- W [18, 24, 30, 35, 40]
- X [15, 20, 26, 30, 37]
- Y [20, 26, 31, 37, 42]
- Z [21, 27, 33, 38, 43]

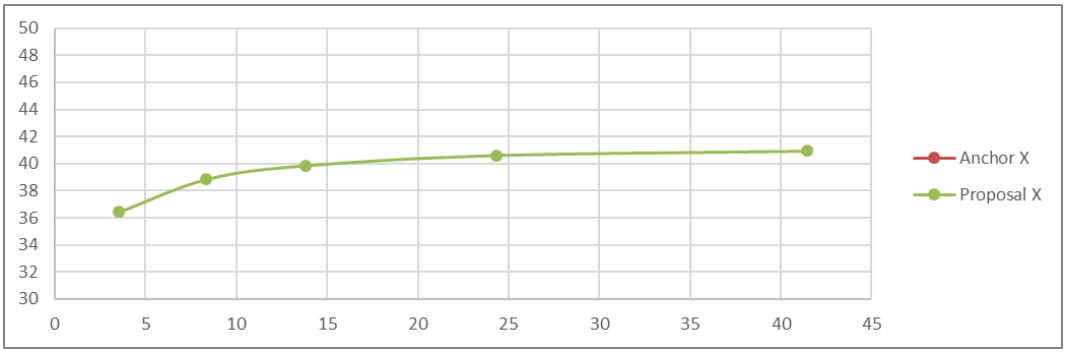
According to objective results (IV-PSNR vs. bitrate figures below), MV-HEVC provides worse quality than the MIV anchor (with HM – obtained in EE2.2) in A97 configuration for all tested sequences. For sequence X the calculation of PSNR for MV-HEVC-encoded data was not possible, as virtual views from RVS 4.0 had some values of luminance greater than the 10-bit range.

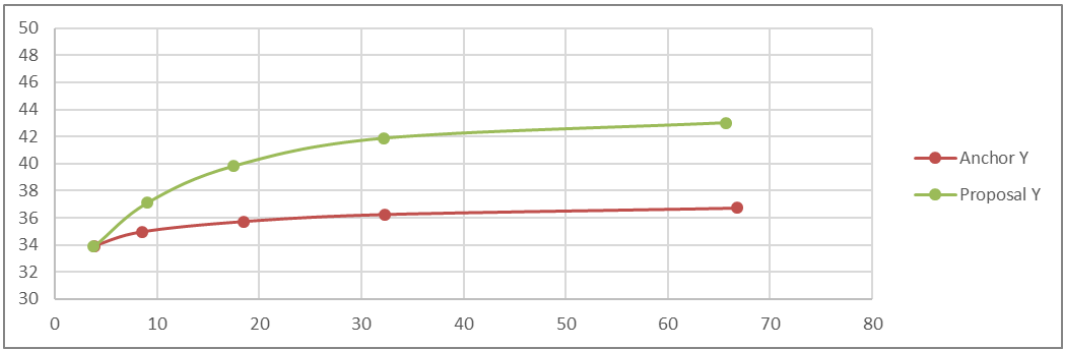


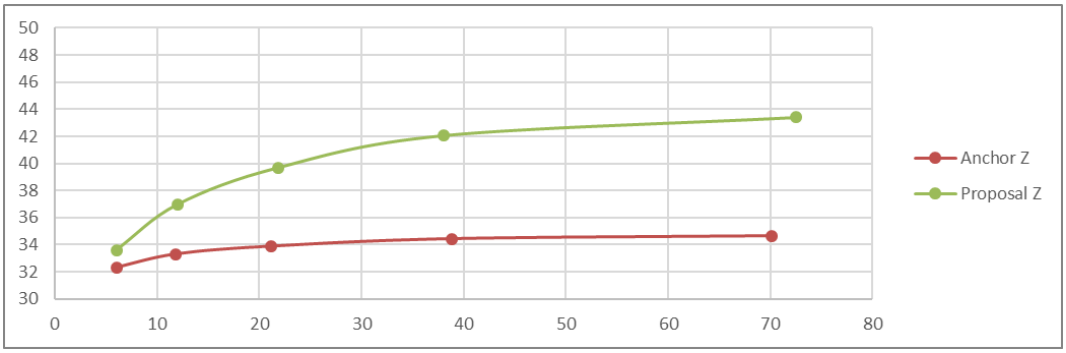





50
48
46
44
42
40
38
36
34
32
30
0
5
10
15
20
25
30
35
40
Anchor W
Proposal W


50
48
46
44
42
40
38
36
34
32
30
0
5
10
15
20
25
30
35
40
45
Anchor X
Proposal X


50
48
46
44
42
40
38
36
34
32
30
0
10
20
30
40
50
60
70
80
Anchor Y
Proposal Y


50
48
46
44
42
40
38
36
34
32
30
0
10
20
30
40
50
60
70
80
Anchor Z
Proposal Z

Recommendations:

InterDigital:

- Use provided QPs and coding results in VT.

PUT:

- Perform remote expert viewing using provided posetraces.
- Fix an error in RVS 4.0 that is causing the luminance values to be above the 10-bit range.

EE3: coding and rendering of non-Lambertian content

Owner: Sarah Fachada (ULB)

Description: RVS4.0 was designed to render features visible on non-Lambertian surfaces. Objective results show superior performance compared objectively and subjectively on Magritte sequence m57103. Currently, this tool is not embedded in TMIV. The process for this experiment will be as follows:

- Anchors is RVS 3.1 using 1 texture + 1 depth (estimated using IVDE v4.1).
- Results for Mirror sequence should be recomputed since wrong depth maps were used

The non-Lambertian scene has been rendered using RVS3.1 CPU/GPU+ground truth depth map, RV3.1 CPU/GPU+IVDE depth maps and RVS4.0 GPU+"multidepth" non-Lambertian maps.

Participants:

Organization	Contact
ULB	Sarah Fachada
ETRI-MC	Gun Bang
ZJU	Sicheng Li
PUT (depth maps)	Dawid Mieloch

Cross-check: The objective metrics for all datasets were computed by Sarah Fachada. Gun Bang computed them from Mirror and Cadillac, Sicheng Li for Magritte T and Magritte M. The crosscheck was successful.

Results: Depending on the number of input images (4 or 6 for Mirror and Cadillac, 4 or 9 for Magritte) the kind of non-Lambertian object (Cadillac: semi reflective, Mirror: planar mirror, Magritte: fully refractive/reflective sphere), the best performing method varies.

Mirror:

4 input	4 - GT -CPU		4 -GT -GPU - 3.1		4 - IVDE - CPU		4 - IVDE - GPU - 3.1		4 - GTMULTI	
	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR
MEAN wo ref	25.4723182	30.1077	25.2084909	29.9370818	26.3096909	32.1245636	26.4154	32.0625545	25.5440818	31.3577727
6 input	6 - GT - CPU		6 -GT - GPU -3.1		6 - IVDE - CPU		6 - IVDE - GPU -3.1		6 - GTMULTI	
	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR
MEAN wo ref	26.2535889	30.9321667	25.9492222	30.7559444	26.7636889	32.4053222	26.8625778	32.3341889	26.1996222	32.1388667

Cadillac:

4 input	4 - GT -CPU		4 -GT -GPU - 3.1		4 - IVDE - CPU		4 - IVDE - GPU - 3.1		4 - GTMULTI	
	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR
MEAN wo ref	28.3463	34.1591818	26.7764545	33.2588	25.8985545	30.9776727	24.5596091	29.1221636	26.8383909	33.2129182
6 input	6 - GT - CPU		6 -GT - GPU -3.1		6 - IVDE - CPU		6 - IVDE - GPU -3.1		6 - GTMULTI	
	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR
MEAN wo ref	29.1765222	34.7132111	27.4036667	33.8727667	26.4063333	31.4459667	25.0055222	29.4919444	27.8394889	34.3237333

Magritte T:

4 input	4 - GT -CPU		4 -GT -GPU - 3.1		4 - IVDE - CPU		4 -IVDE -GPU - 3.1		4 - GTMULTI		4 - IVDEMULTI	
	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR
MEAN	29.0516524	35.5889333	28.8433714	35.6052333	23.2111524	30.245381	23.2694048	30.4171762	29.0384048	37.1795333	23.4323238	30.7769095
9 input	9 - GT - CPU		9 -GT -GPU - 3.1		9 - IVDE - CPU		9 -IVDE -GPU - 3.1		9 - GTMULTI		9 - IVDEMULTI	
	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR
MEAN	30.6594381	36.9161952	30.4227381	36.9775762	24.928819	31.5932762	25.1390476	31.7960429	31.1201048	40.3445095	28.3382	37.6946571

Magritte M:

4 - GT -CPU		4 -GT -GPU - 3.1		4 - IVDE - CPU		4 -IVDE -GPU - 3.1		4 - GTMULTI		4 - IVDEMULTI	
PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR
29.4429667	35.2033619	29.1876143	35.2557381	23.8928381	30.9074571	23.9319286	31.0895429	29.8588714	36.8367667	23.8940667	30.9695905
9 - GT - CPU		9 -GT -GPU - 3.1		9 - IVDE - CPU		9 -IVDE -GPU - 3.1		9 - GTMULTI		9 - IVDEMULTI	
PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR	PSNR	IVPSNR
31.2963619	36.7406381	30.9967381	36.8369381	25.9735857	32.6860143	26.2089762	32.9606667	33.7610333	41.6396524	29.6875048	38.0683429

Recommendations:

(ULB) Provided the results of this experiment, we recommend to estimate the depth maps when the object cannot be considered as Lambertian. With enough input views, multidepth should be computed.

Explorations on new tools is essential to better handle occlusions and create multidepth for datasets with low number of input images or large baselines and continue the EE when the tools are ready.

EE-4: Results of LCEVC in TMIV

Owner: Lorenzo Ciccarelli (V-Nova)

Description:

Experiment EE4 proposes to test the coding efficiency of MIV views using the VVC Test Model with a multi-layer profile to compress the material before being encoded by the MIV framework. In this experiment

the aim is to use LCEVC to compress the atlases generated by the MIV encoder to evaluate the coding efficiency and the encoding runtime speed-up provided by LCEVC compared to the current solution.

Participants:

Philips (@bartkroong), PUT (@dmieloch), ZJU (@SichengLi)

Cross-check:

The experiment results are only partial. Cross-check are suggested to be carried out at the next round.

Results:

(Task 1) Generation of the anchor

Anchors have been identified using the results of [WG04 / N0148]. Along with the metrics summarized in the attached template intermediate files of the VvenC have been provided to allow LCEVC calibration.

(Task 2) Extend TMIV to output 14bit

LCEVC can encode up to 14bit bit-depth. For this reason the TMIV has been modified to provide 14bit geometries as intermediate file to be encoded with LCEVC. The TMIV version used to generate the geometries is the 11.1.

(Task 3) LCEVC configuration

LTM 5.4 (LCEVC Test Model 5.4) has been used to encode the material and define the best configuration to use for both the textures and geometries.

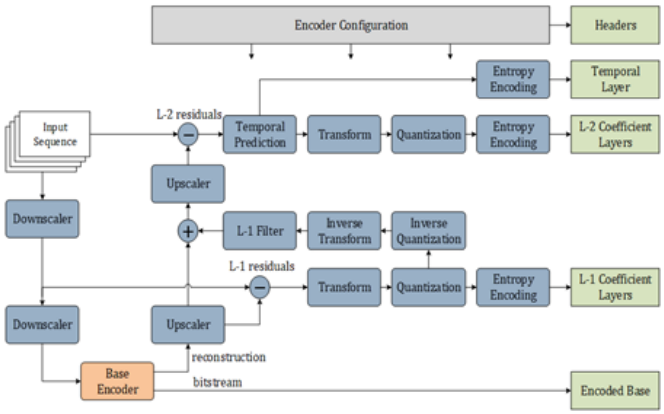


Figure 1 - LCEVC Encoder

As described in Figure 1 LCEVC can use any encoder as base encoder.

The following steps have been followed to find the optimal configuration for the LTM:

1. VVenC 0.3.1.0 has been used to generate the base encoder bitstreams.
2. Calibration experiments have been carried out using A97 class. (A, B, D, E, J, N, O, P, R).
3. For each mandatory sequence half resolution geometries and textures has been created using different types of downsamplers.
4. The material in 3 has been used to produce the VVenC bitstreams as based encoder bitstream to pass to the LCEVC encoder. (note that 14 bit geometries has been converted to 10bit before being used in VVenC).
5. VVenC encoder has been used in slow and slower pre-sets to generate multiple QPs bitstreams. An Excel spreadsheet summarising the information for each bitstream generated has been created. The file has been attached to the input contribution m58968 and it contains information about downsampler type, QP, size in bytes, percentage bytes compared to each anchor QP point for each bitstream.
6. The material generated has been analysed to find out the best QP to use as base bistreams compared to each anchor QP
7. Several combination of LCEVC tools has been tested
8. Information from 6 and 7 has been used in order to maximise the resulting PSNR, VMAF of each QP point when compared to the same metric for the anchors. A visual inspection has been carried out to confirm the base QP and tools set selected.

Table 1 is describing some the best combination found for both geometries and textures.

Tools	**Geometry**	**Texture**
Base bitrate vs Anchor bitrate	Between 50%-60% for all bitrate	Between 65% and 90% depending on the bitrate
Final bitdepth	14bit	10bit
Downsampler	Area downsampler	Lanczos downsampler
Upsampler	Nearest Upsampler	Modified Cubic/custom
Transform	2x2	2x2/4x4
Quantization matrix	Disabled	Default
Temporal Step with multiplier	Always maximum	Depending on the base QP
Predicted residual	Enabled	Enabled
U and V component residual	Disabled	Enabled

Table 2 - LCEVC Tool configuration

The configuration above and the relation between the target rate and the base QP are reported in the m58968 input contribution. The json files used to configure the LTM will be attached to allow cross-checks.

(Task 4) Generation of the LCEVC bitstreams

Given the very peculiar nature of the content the initial configuration selected to carry out the experiment has been used only for A97 mandatory sequences. For each texture, geometry and bitrate point an LCEVC bitstream has been generated. To allow the selection of the quantization parameter (stepwidth) for each level of enhancement an hunting algorithm has been used to match a the best precision the anchors rates.

(Task 5) Comparison between anchor and target

In order to verify the chosen configuration each LCEVC bitstream has been decoded using LTM decoder and then passed as out-of-band material to the TMIV. The version used for the TMIV was the 11.1. All the views and 3 poses for each rate has been generated.

PSNR and IV-PSNR values has been calculate in order to compare to the anchor

After having synthetised the view and the post-traces 9t was noted that TMIV 11.1 had some problem so it has been recommended to repeat this part of the experiment using TMIV 12.0, however the process cannot be finished in time for this meeting.

The recommendation is to extend also the metric to MS-SSIM and VMAF that needs to be calculated also for the anchor.

To show some preliminary results some visual inspections have been carried out. Following just an example of few pose traces generated. The following picture are pose 1 and 2 of seq A at the lowest bitrate.

LCEVC	VVENC
	

Table 3 - Pose 2 lowest bitrate (LCEVC left / Anchor right)

LCEVC	VVENC
	

Table 4 - Pose 1 lowest bitrate (LCEVC left / Anchor right)

Recommendations:

Given that the test has been carried out only on A97 sequence and that the results were based on the TMIV 11.1 the recommendation is to continue this experiment to cover the following :

- Extend the generation of the LCEVC intermediate files to the A17 and V17 classes.
- Generate A97, A17 and V17 anchor metrics including VMAF and MS-SSIM
- Repeat the test using TMIV 12.0

EE-5: Decoder-side depth estimation

Owner: Adrian Dziembowski (PUT)

Description:

EE5.5: the goal is to 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.

EE5.6: the goal is to 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.

EE5.7: the goal is to test whether it is better to filter the textures before or after feature extraction.

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

Cross-check of EE5.5:

The crosscheck was not performed, because the configuration files were not provided in time.

Results of EE5.5:

EE5.5-1 (GA SEI for all views):

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 [%]
Painter	D	13.8%	10.6%	9.8%	6.9%	282.7%	101.6%	49.0%	7.15	7.99	11.7%	6.42	7.03	9.6%
Frog	E	1.1%	5.1%	5.9%	8.5%	215.3%	92.3%	13.9%	7.50	7.55	0.8%	7.31	7.65	4.6%
Kitchen	J	20.6%	19.7%	8.2%	14.9%	796.1%	104.7%	179.2%	12.74	12.65	-0.7%	12.48	11.89	-4.7%
Carpark	P	-18.5%	-7.8%	-23.9%	-13.1%	211.1%	73.9%	40.9%	10.23	9.70	-5.1%	8.19	7.38	-9.9%
Fan	O	5.3%	8.6%	0.4%	5.7%	357.5%	81.1%	13.9%	10.99	10.56	-3.9%	10.11	9.31	-7.9%
Group	R	---	---	---	---	553.1%	98.4%	28.4%	22.51	16.81	-25.3%	23.48	16.85	-28.3%
MIV		**---**	**---**	**---**	**---**	**402.6%**	**92.0%**	**54.2%**	**11.85**	**10.88**	**-3.8%**	**11.33**	**10.02**	**-6.1%**

EE5.5-2 (GA SEI for views in first atlas, no recursion):

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 [%]
Painter	D	449.8%	169.2%	275.2%	141.4%	356.0%	107.6%	76.3%	7.15	12.79	78.9%	6.42	12.09	88.3%
Frog	E	0.8%	5.0%	8.6%	10.5%	299.4%	109.9%	57.0%	7.50	7.54	0.5%	7.31	7.66	4.7%
Kitchen	J	-19.9%	-6.1%	-24.2%	-7.8%	972.8%	120.8%	60.2%	12.74	12.08	-5.1%	12.48	10.93	-12.4%
Carpark	P	-11.6%	-2.6%	-30.2%	-18.1%	265.6%	128.2%	98.0%	10.23	9.73	-4.9%	8.19	7.18	-12.3%
Fan	O	-2.9%	3.6%	-4.2%	3.2%	440.3%	121.3%	58.1%	10.99	10.45	-4.9%	10.11	9.26	-8.4%
Group	R	---	---	---	---	659.6%	118.4%	61.3%	22.51	16.62	-26.2%	23.48	16.97	-27.7%
MIV		**---**	**---**	**---**	**---**	**498.9%**	**117.7%**	**68.5%**	**11.85**	**11.54**	**6.4%**	**11.33**	**10.68**	**5.4%**

- there is a bug in IVDE, which significantly lowers the quality for SD,
- initial grid size for EE5.5-2 (32x32) was too small, thus high quantization has to be used in order to fit within the 1Mbps feature metadata limit.

EE5.5-3 (GA SEI for views in first atlas, recursion): no results yet.

Cross-check of EE5.6 (PUT/Tencent):

- EE5.6-1: perfect match,
- EE5.6-2: perfect match except for SP (exact bitrates, max PSNR diff: 0.3 dB, avg diff: 0.03 dB),
- crosscheck performed for mandatory content.

Results of EE5.6:

EE5.6-1 (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 [%]
Painter	D	15.1%	23.4%	2.3%	13.3%	342.8%	129.2%	52.5%	7.15	6.85	-4.2%	6.42	4.73	-26.2%
Frog	E	15.7%	21.0%	25.1%	26.7%	279.3%	92.8%	41.9%	7.50	7.40	-1.3%	7.31	7.75	5.9%
Kitchen	J	3.8%	0.2%	-5.4%	-3.0%	891.5%	80.2%	49.9%	12.74	11.56	-9.2%	12.48	11.22	-10.1%
Carpark	P	83.0%	70.6%	39.5%	43.3%	270.6%	93.9%	54.9%	10.23	10.05	-1.8%	8.19	7.98	-2.5%
Fan	O	34.6%	69.3%	30.3%	66.2%	443.7%	95.7%	61.2%	10.99	10.11	-8.1%	10.11	9.04	-10.6%
Group	R	---	---	---	---	627.8%	81.3%	47.3%	22.51	19.39	-13.9%	23.48	21.05	-10.3%
MIV		---	---	---	---	**475.9%**	**95.5%**	**51.3%**	**11.85**	**10.89**	**-6.4%**	**11.33**	**10.30**	**-9.0%**
Optional content - Proposal vs. Low/High-bitrate Anchors														
ClassroomVideo	A	-69.3%	-39.4%	-42.2%	-23.3%	1736.0%	90.2%	41.9%	5.69	4.85	-14.8%	4.06	3.00	-26.1%
Museum	B	63.1%	30.6%	12.7%	7.9%	1755.6%	112.3%	50.2%	9.27	10.40	12.2%	6.46	7.79	20.6%
Hijack	C	---	---	---	---	753.9%	93.1%	54.4%	22.25	25.62	15.2%	20.97	24.14	15.1%
Mirror	I	15.5%	27.7%	-10.2%	17.7%	279.3%	113.5%	57.6%	13.10	12.92	-1.3%	12.96	11.47	-11.5%
Cadillac	G	3.3%	19.5%	6.1%	18.4%	562.1%	145.0%	54.1%	14.49	13.91	-4.0%	14.56	14.26	-2.0%
Fencing	L	-70.0%	-21.4%	-10.1%	13.0%	358.9%	113.1%	52.7%	12.90	12.88	-0.1%	9.18	8.95	-2.4%
Chess	N	---	---	---	---	491.7%	106.8%	61.2%	24.33	28.44	16.9%	23.08	26.90	16.6%
ChessPieces	Q	---	---	---	---	552.0%	110.9%	60.9%	27.96	30.65	9.6%	26.03	29.02	11.5%
Hall	T	---	---	601.6%	485.0%	374.2%	138.6%	48.6%	15.86	17.62	11.1%	13.16	15.80	20.1%
Street	U	5.2%	10.6%	15.1%	18.5%	265.9%	139.3%	50.1%	7.07	7.02	-0.7%	4.91	4.68	-4.6%
MIV		---	---	---	---	**713.0%**	**116.3%**	**53.2%**	**15.29**	**16.43**	**4.4%**	**13.54**	**14.60**	**3.7%**

EE5.6-2 (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 [%]
Painter	D	37.9%	55.6%	20.3%	40.1%	276.8%	127.2%	32.2%	7.15	6.51	-8.9%	6.42	4.01	-37.6%
Frog	E	35.3%	44.2%	42.7%	48.7%	223.1%	122.0%	21.4%	7.50	7.48	-0.3%	7.31	7.79	6.5%
Kitchen	J	-19.8%	-8.2%	-26.7%	-11.3%	783.4%	79.0%	40.4%	12.74	10.51	-17.5%	12.48	9.58	-23.2%
Carpark	P	107.7%	109.8%	62.2%	69.9%	216.1%	122.9%	25.0%	10.23	10.37	1.4%	8.19	8.05	-1.6%
Fan	O	84.4%	151.0%	64.5%	134.3%	372.8%	123.8%	53.4%	10.99	9.95	-9.5%	10.11	8.62	-14.7%
Group	R	---	---	---	---	556.1%	81.2%	40.0%	22.51	19.15	-14.9%	23.48	20.72	-11.7%
MIV		---	---	---	---	**404.7%**	**109.3%**	**35.4%**	**11.85**	**10.66**	**-8.3%**	**11.33**	**9.80**	**-13.7%**
Optional content - Proposal vs. Low/High-bitrate Anchors														
ClassroomVideo	A	-48.3%	-3.3%	-16.7%	7.7%	1443.5%	82.5%	30.7%	5.69	6.30	10.7%	4.06	4.68	15.1%
Museum	B	-11.7%	11.4%	-1.7%	13.4%	1538.5%	97.0%	43.6%	9.27	9.57	3.2%	6.46	7.09	9.7%
Hijack	C	---	---	---	---	653.6%	104.7%	44.8%	22.25	26.34	18.4%	20.97	25.13	19.9%
Mirror	I	17.8%	39.0%	-12.7%	22.8%	241.0%	125.5%	41.2%	13.10	12.65	-3.4%	12.96	11.08	-14.6%
Cadillac	G	9.7%	40.4%	4.8%	33.9%	476.8%	155.4%	38.3%	14.49	13.45	-7.2%	14.56	13.59	-6.7%
Fencing	L	-58.9%	1.8%	5.4%	38.3%	273.1%	108.8%	12.9%	12.90	12.89	0.0%	9.18	9.02	-1.7%
Chess	N	---	---	---	---	408.6%	100.8%	57.6%	24.33	26.93	10.7%	23.08	26.10	13.1%
ChessPieces	Q	---	142.2%	---	261.1%	433.2%	130.5%	66.2%	27.96	29.04	3.9%	26.03	27.76	6.6%
Hall	T	820.7%	167.2%	-14.0%	7.0%	304.7%	208.5%	21.1%	15.86	16.18	2.0%	13.16	13.11	-0.4%
Street	U	20.3%	26.9%	27.1%	33.8%	234.2%	163.8%	50.5%	7.07	7.04	-0.4%	4.91	4.71	-4.0%
MIV		---	---	---	---	**600.7%**	**127.7%**	**40.7%**	**15.29**	**16.04**	**3.8%**	**13.54**	**14.23**	**3.7%**

One geometry atlas vs. 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 [%]
Painter	D	12.3%	19.2%	9.0%	16.6%	80.7%	90.8%	61.3%	6.85	6.51	-4.9%	4.73	4.01	-15.4%
Frog	E	16.7%	19.0%	14.1%	17.3%	79.9%	130.0%	51.1%	7.40	7.48	1.0%	7.75	7.79	0.5%
Kitchen	J	-37.6%	-25.8%	-37.6%	-26.7%	87.9%	77.0%	81.0%	11.56	10.51	-9.1%	11.22	9.58	-14.6%
Carpark	P	17.4%	22.4%	15.6%	19.0%	79.9%	128.6%	45.5%	10.05	10.37	3.2%	7.98	8.05	0.9%
Fan	O	34.6%	46.5%	25.0%	39.3%	84.0%	127.4%	87.3%	10.11	9.95	-1.6%	9.04	8.62	-4.6%
Group	R	-75.3%	-70.0%	-72.6%	-62.2%	88.6%	76.1%	84.5%	19.39	19.15	-1.2%	21.05	20.72	-1.6%
MIV		**-5.3%**	**1.9%**	**-7.7%**	**0.5%**	**83.5%**	**105.0%**	**68.4%**	**10.89**	**10.66**	**-2.1%**	**10.30**	**9.80**	**-5.8%**
Optional content - Proposal vs. Low/High-bitrate Anchors														
ClassroomVideo	A	37.9%	22.3%	11.2%	10.8%	83.1%	71.9%	73.2%	4.85	6.30	29.9%	3.00	4.68	55.7%
Museum	B	-47.7%	-27.8%	-28.6%	-15.1%	87.6%	71.5%	86.8%	10.40	9.57	-8.0%	7.79	7.09	-9.0%
Hijack	C	---	---	---	---	86.7%	89.7%	82.3%	25.62	26.34	2.8%	24.14	25.13	4.1%
Mirror	I	2.8%	9.1%	-0.4%	4.5%	86.3%	108.7%	71.4%	12.92	12.65	-2.1%	11.47	11.08	-3.4%
Cadillac	G	2.9%	16.0%	-3.1%	12.3%	84.8%	105.4%	70.8%	13.91	13.45	-3.4%	14.26	13.59	-4.7%
Fencing	L	32.5%	27.7%	15.7%	21.6%	76.1%	95.2%	24.4%	12.88	12.89	0.1%	8.95	9.02	0.7%
Chess	N	---	---	---	---	83.1%	81.5%	94.1%	28.44	26.93	-5.3%	26.90	26.10	-3.0%
ChessPieces	Q	---	---	---	---	78.5%	102.3%	108.8%	30.65	29.04	-5.2%	29.02	27.76	-4.3%
Hall	T	---	---	---	---	81.4%	149.1%	43.5%	17.62	16.18	-8.2%	15.80	13.11	-17.1%
Street	U	13.6%	14.6%	10.2%	12.6%	88.1%	115.8%	100.7%	7.02	7.04	0.3%	4.68	4.71	0.7%
MIV		**---**	**---**	**---**	**---**	**83.6%**	**99.1%**	**75.6%**	**16.43**	**16.04**	**0.1%**	**14.60**	**14.23**	**2.0%**

Comments from PUT:

- the "AutomaticDepthRange" parameter has to be set to false when we send input depth maps with already known ZRange,
- the total number of atlases in EE5.6-2 is five (3 attribute + 2 geometry atlases), however, the geometry atlases have reduced resolution, so they could be potentially packed,
- sending of input depth maps allows to significantly reduce the decoding time (by 50% when 1 geometry atlas is available, and by 60% when two atlases are sent),
- for a majority of perspective content, the objective BD rates are worse because of similar quality and increase of the bitrate,
- subjectively, the posetraces for the approach with input depth assistance are more stable and consistent, than for the G17 anchor,
- for SN, SC, SQ, the bug in IVDE reduces the efficiency of the approach with input depth assistance,
- the same bug probably lowers the quality for SB, however, the current results are already much better than the G17 anchor,
- approach with two geometry atlases seems to be more efficient when there are more views, or the cameras captured the scene from very different angles (i.e., SR).

Comments from Tencent:

- Test 1 and test 2 don't bring gain over G17 anchor.

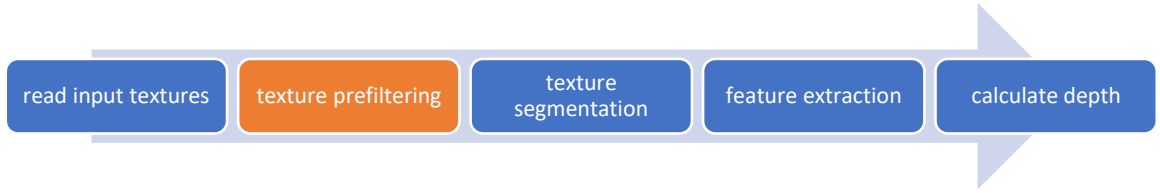
Cross-check of EE5.7 (PUT/ETRI-IM):

- EE5.7-1: no crosscheck needed (EE5.7-1 is the G17 anchor),
- EE5.7-2: perfect match.

Results of EE5.7:

EE5.7-2 (filtering of textures after feature extraction):

G17 anchor/EE5.7-1:



EE5.7-2:

Objective comparison of G17 anchor vs. EE5.7-2:

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
Painter	D	-2.1%	-0.8%	-1.2%	-0.5%
Frog	E	-0.0%	-0.0%	0.1%	0.2%
Kitchen	J	-1.8%	-2.2%	-0.3%	-1.6%
Carpark	P	2.2%	-0.4%	4.1%	1.8%
Fan	O	0.5%	-0.0%	4.7%	2.6%
Group	R	-22.2%	-15.7%	-23.0%	-13.5%
MIV		**-3.9%**	**-3.2%**	**-2.6%**	**-1.8%**

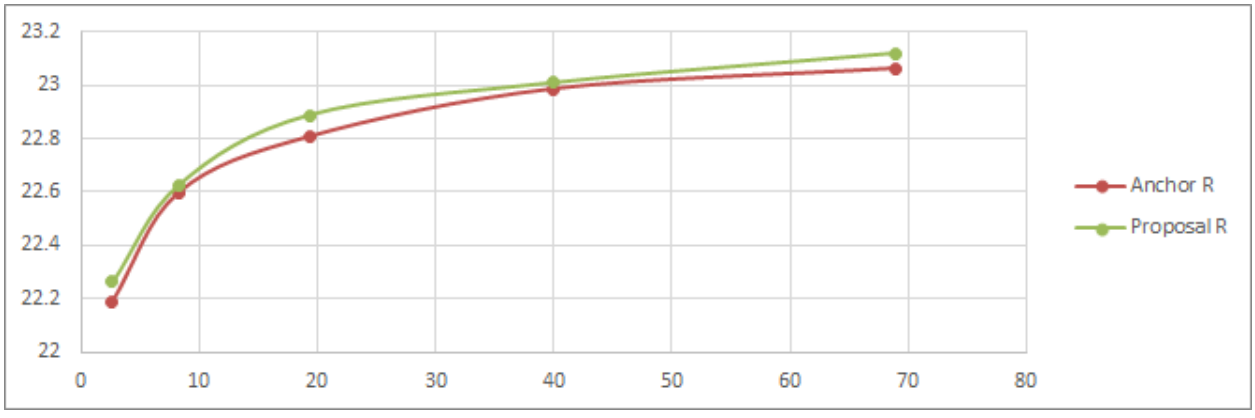
Runtime ratio (%)		
Atlas encoding	Video encoding	Decoding & Rendering
97.9%	103.6%	80.1%
105.9%	94.4%	82.7%
133.1%	111.1%	99.5%
71.2%	92.3%	88.3%
74.2%	115.4%	106.1%
96.8%	110.3%	81.1%
96.5%	**104.5%**	**89.6%**

Max delta Y-PSNR [dB]		
MIV DSDE	########	Difference [%]
7.15	6.92	-3.2%
7.50	7.47	-0.4%
12.74	13.06	2.5%
10.23	10.21	-0.2%
10.99	10.84	-1.4%
22.51	22.45	-0.3%
11.85	**11.82**	**-0.5%**

Max delta IV-PSNR [dB]		
MIV DSDE	########	Difference [%]
6.42	6.49	1.1%
7.31	7.29	-0.4%
12.48	13.07	4.8%
8.19	8.26	0.9%
10.11	10.21	1.0%
23.48	23.39	-0.4%
11.33	**11.45**	**1.2%**

Comments from PUT:

- for 5 of 6 sequences, the differences are negligible, and we cannot say, that one approach is better than the other,
- for SR we can see a huge BD difference both for PSNR and IV-PSNR, and the curve for EE5.7-2 is noticeably higher, than for the G17 anchor:



- nevertheless, we cannot say, that the subjective quality for SR is better, as the quality for some views grew up from 17.3 to 17.6, which is still ridiculously low; the synthesized views look different, but equally bad:

SR, v2, EE5.7-1 (G17 anchor), 17.3 dB:	SR, v2, EE5.7-2, 17.6 dB:
	

Comments from ETRI-IM:

- Even if a fair amount of gain can be achieved at sequence R, the average gain across all tested sequences seems too small to change the order of DSDE anchor generation

Recommendations:

- EE5.5
  - PUT
    - repeat the experiment if the IVDE bug will be fixed, and if proper configuration files will be provided,
  - ETRI-IM
    - Keep EE 5.5 since it was not carried out during this meeting cycle,
- EE5.6
  - PUT
    - repeat the experiment when the bug in IVDE will be resolved,
  - Tencent
    - keep investigating other strategies of depth/texture allocation,
- EE5.7
  - PUT
    - no change in the default order and keep the feature extraction on prefiltered textures,
  - ETRI-IM
    - Maintain the current DSDE anchor generation workflow, which is doing feature extraction after texture prefiltering.

EE6: Correlation of objective and subjective evaluations for future MIV

Owner: Joel Jung (Tencent)

Participants: Franck Thudor (Interdigital), Vinod Kumar Malamal Vadakital (Nokia), Bart Kroon (Philips), Sicheng Li (ZJU), Joel Jung (Tencent)

Description:

EE6.1: the goal is to evaluate how much objective metrics match with MOS on pose traces, when the reference is the ground-truth or the best-reference.

EE6.2: the goal is to evaluate how much objective metrics match with MOS source view positions, when the reference is the ground-truth captured content.

Cross-check of 6.2: the objective metrics were computed by Franck Thudor (Interdigital) and Joel Jung (Tencent) and successfully cross-checked. The correlations were computed by Bart Kroon (Philips) and Joel Jung (Tencent) and successfully cross-checked.

Results of 6.1:

Ground truth results were made available only for ClassroomVideo. From this result it has been observed that the best reference significantly differs from the ground-truth reference. It has been concluded that the best-reference cannot be used to compute full-reference objective metrics. As a consequence, the viewing session has not been performed.

Results of 6.2:

Results of objective metrics:

	VMAF	PSNR	IV-PSNR	SSIM	WS-PSNR
Chess RP2 v3	81.96138	35.41672	46.8035	0.983899	34.4021
Chess RP4 v3	77.03951	34.25814	44.2528	0.979272	33.2058
Chess RP2 v6	89.02366	39.92507	50.7819	0.995535	38.6601
Chess RP4 v6	83.15557	37.27157	46.0482	0.991758	36.1004
ClassroomVideo RP2 v3	88.96994	35.42015	44.9711	0.994302	35.4172
Classroom Video RP4 v3	80.36913	33.75792	41.6758	0.9819	33.6112
ClassroomVideo RP2 v11	86.3749	34.64424	44.476	0.992729	34.6916
ClassroomVideo RP4 v11	78.18641	33.23045	41.3726	0.980851	33.0983
Frog RP2 v6	89.52923	31.77308	41.5073	0.979056	31.7476
Frog RP4 v6	75.78866	28.94353	37.3141	0.941935	28.918
Frog RP2 v10	81.65114	29.76867	39.4608	0.96747	29.7432
Frog RP4 v10	71.65907	27.83585	36.3288	0.93319	27.8103
Painter RP2 v2	90.28133	38.03019	46.3563	0.990799	38.0047
Painter RP4 v2	71.65481	33.17622	40.22	0.949013	33.1507
Painter RP2 v8	87.87954	35.37033	43.7736	0.986006	35.3448
Painter RP4 v8	70.26109	31.63682	38.7801	0.935428	31.6113
Barn RP2 v1	79.33873	31.3525	41.5456	0.978384	31.327

Barn RP4 v1	67.68332	28.4904	37.2704	0.943602	28.4649
Barn RP2 v13	78.70313	31.2882	41.4842	0.977976	31.2627
Barn RP4 v13	66.9791	28.5732	37.3012	0.942989	28.5477
Museum RP2 v1	82.32553	31.22114	40.6212	0.993392	31.547
Museum RP4 v1	61.49144	26.81629	35.0543	0.957895	27.3206
Museum RP2 v8	74.7497	28.60481	39.8452	0.988086	29.4968
Museum RP4 v8	56.048	25.68788	34.3825	0.949495	26.3056
CarPark RP2 v1	83.22891	34.37113	42.5197	0.982719	34.3456
CarPark RP4 v1	77.13061	32.14094	39.468	0.956889	32.1154
Carpark RP2 v7	83.7333	34.90315	43.6839	0.982431	34.8776
CarPark RP4 v7	78.47068	32.7987	40.229	0.958377	32.7732

Results of correlation computations:

	Pearson correlation	Spearman correlation	Kendal correlation	RMSE (9 grade scale)
VMAF	0.76	0.8	0.61	1.11
PSNR	0.54	0.53	0.38	1.43
SSIM	0.87	0.88	0.74	0.84
IV-PSNR	0.64	0.67	0.46	0.30
WS-PSNR	0.57	0.56	0.4	1.40

Recommendations:

Tencent:

From the correlations results, we believe that SSIM can be safely used to assess MIV content on source view positions for the tested sequences and rate points. Our recommendations are:

- To consider other sequences, other rate points, and confirm further this observation with a new round of subjective tests.
- To apply similar tests on pose traces, using ground truth references, to check is a similar conclusion can be made for pose traces.