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/WG4 m55712
October 2020, Online

Source Tencent, Nokia, Poznan University of Technology
Status Input document
Title Report of the Exploration Experiments on Future MPEG Immersive Video
Author Joel Jung, Vinod Kumar Malamal Vadakital, Dawid Mieloch

Abstract

This document summarizes the results obtained by the different organizations participating to the EE on Future MPEG Immersive Video. This EE, described in [WG04N0010], evaluates in a collaborative process, the coding of TMIV generated attribute and geometry atlases with VVC, x265 and MV-HEVC, instead of HEVC, a frame packing method with VVC, and decoder-derived depth.

Exploration Experiment EE1: VVC-related

EE1.a: TMIV + VVC anchor generation

This exploration experiment was designed to generate MIV anchors using the VTM codec implementation of VVC standard. VTM version 10.1 with random-access + Class-F configuration was used for simulations.

Participants
The workload of conducting this experiment is split, based on sequences, among the participants in the following way:

Seq	**SA**	**SB**	**SD**	**SE**	**SJ**	**SN**	**SO**	**SP**	**SR**
Tester 1	Philips		Nokia		Philips	Nokia	Oppo		Philips
Tester 2	ZJU	Nokia	ZJU	Intel	PUT	Oppo	ZJU	PUT	PUT
Tester 3	Tencent	Tencent	Tencent	Tencent	Tencent	Tencent	Tencent	Tencent	Tencent

Results of the cross-check:

- Cross check of sequences SA, SD, SE, SN, SP showed differences; the differences from objective evaluation seemed to be minor.
- Cross check of sequences SB, SJ, SO were exact.

Tester 3 matches perfectly tester 1 for SA, SD, SJ, SN and matches perfectly tester 2 for SB, SE, SJ. Tester 3 results are wrong for SP and SR.
Results:

The table below reports the results as produced by the experimenters. It is confirmed that VVC brings significant gains in the coding of MIV content. However, the differences in the produced results among testers is a cause for concern which requires further investigation.

Testers 1

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	
ClassroomVid	SA	-46.7%	-40.7%	1.51	-47.5%	-30.1%	-31.9%	-32.1%	Philips
Museum	SB	-25.2%	-26.0%	16.27	#VALUE!	#VALUE!	-20.9%	-21.4%	Tencent
Fan	SO	-39.8%	-37.2%	8.01	-44.0%	-42.8%	-35.1%	-33.0%	Oppo
Kitchen	SJ	-20.2%	-26.0%	16.37	-28.3%	-31.7%	-25.0%	-23.7%	Philips
Painter	SD	-24.8%	-23.7%	7.93	-26.8%	-26.3%	-16.7%	-17.2%	Nokia
Frog	SE	-33.2%	-32.0%	6.55	#VALUE!	#VALUE!	-27.6%	-27.2%	Tencent
Carpark	SP	-29.3%	-29.2%	7.17	#VALUE!	#VALUE!	-16.5%	-20.5%	Tencent
Chess	SN	-32.8%	-32.2%	14.71	-37.4%	-35.7%	-34.1%	-31.7%	Nokia
Group	SR	-36.8%	-34.5%	12.19	-39.4%	-37.6%	-27.6%	-25.7%	Philips

Tester 2

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	
ClassroomVid	SA	-46.5%	-40.8%	1.51	-49.1%	-31.7%	-32.0%	-32.4%	ZJU
Museum	SB	-25.2%	-26.0%	16.27	-26.2%	-28.5%	-20.9%	-21.4%	Nokia
Fan	SO	-39.8%	-37.2%	8.01	-44.0%	-42.8%	-35.1%	-33.0%	ZJU
Kitchen	SJ	-20.2%	-26.0%	16.37	-28.3%	-31.7%	-25.0%	-23.7%	PUT
Painter	SD	-27.1%	-24.7%	7.93	-28.4%	-27.8%	-17.4%	-17.2%	ZJU
Frog	SE	-33.2%	-32.0%	6.55	-37.1%	-35.3%	-27.6%	-27.2%	Intel
Carpark	SP								PUT
Chess	SN	-34.2%	-32.2%	23.49	-46.3%	-37.4%	-34.6%	-31.7%	Oppo
Group	SR	-36.8%	-34.5%	12.19	-39.4%	-37.6%	-27.6%	-25.7%	PUT

Tester 3 (Tencent)

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	Match
ClassroomVid	SA	-46.7%	-40.7%	1.51	#VALUE!	#VALUE!	-31.9%	-32.1%	T1
Museum	SB	-25.2%	-26.0%	16.27	#VALUE!	#VALUE!	-20.9%	-21.4%	T2
Fan	SO	-40.9%	-38.7%	8.01	#VALUE!	#VALUE!	-35.5%	-33.5%	
Kitchen	SJ	-20.2%	-26.0%	16.37	#VALUE!	#VALUE!	-25.0%	-23.7%	T1 & T2
Painter	SD	-24.8%	-23.7%	7.93	#VALUE!	#VALUE!	-16.7%	-17.2%	T1
Frog	SE	-33.2%	-32.0%	6.55	#VALUE!	#VALUE!	-27.6%	-27.2%	T2
~~Carpark~~	~~SP~~	~~-29.3%~~	~~-29.2%~~	~~7.17~~	~~#VALUE!~~	~~#VALUE!~~	~~-16.5%~~	~~-20.5%~~	
Chess	SN	-32.8%	-32.2%	14.71	#VALUE!	#VALUE!	-34.1%	-31.7%	T1
~~Group~~	~~SR~~	~~-37.0%~~	~~-35.6%~~	~~11.74~~	~~#VALUE!~~	~~#VALUE!~~	~~-28.6%~~	~~-26.3%~~	~~Strange!!~~

Summary of recommendations from experimenters:

OPPO:

- Using VVC to encode and decode can indeed improve subjective quality compared to HEVC
- However, VTM spent more than 10 times longer on encoding time.
- Carefully weigh the gains and costs of performance by the different video codecs.

Tencent:

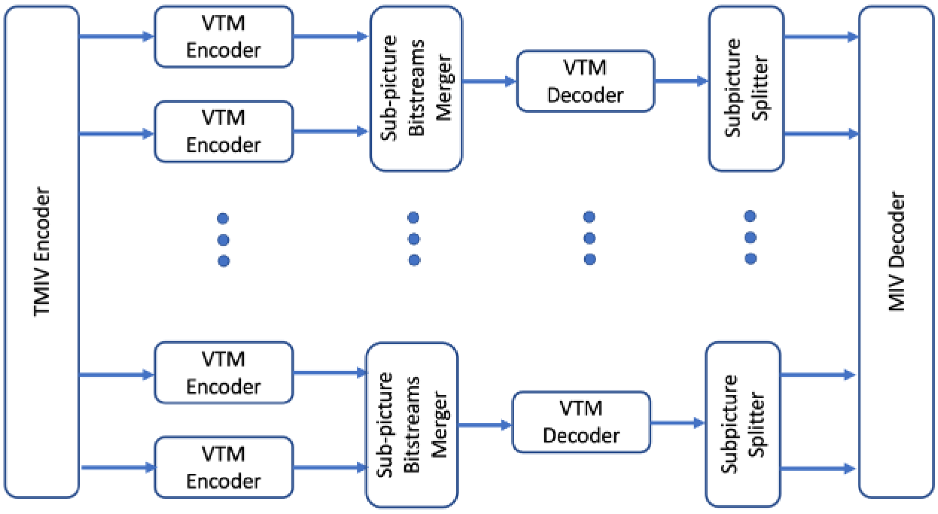
- Considering the large gain and high quality provided by VVC, Tencent recommends adopting VVC as the main anchor for the CTCs, as soon as possible.
- A shortcoming of VTM is the long runtimes for encoding (e.g. SE, QP1, 17 frames take 12 hours).
- To improve the speed of VVC experiments, the group could test through EEs, the usage of faster VVC implementations, or faster encoder configurations for VTM.

Philips:

- The gains produced by VTM 10.1 are in line with results produced by previous EEs.
- By accident, EE1.a was run with the VTM version and configuration of the previous EE. It appears that the VTM versions provide an almost identical performance.

EE1.b: frame packing

In this experiment, VVC is used to encode the attribute and geometry atlases separately. The configuration for VVC encoding and sub-picture merging was made available to the experimenters, by Nokia, at the T1 deadline (20/11/2020). The two bitstreams are merged into a single bitstream, which is sent to the MIV decoder [m54274]. A python script, also provided to the experimenters by Nokia, was used to unpack the packed decoded pictures.



Participants

The workload of conducting this experiment is split, based on sequences, among the participants in the following way:

Seq	**SA**	**SB**	**SD**	**SE**	**SJ**	**SN**	**SO**	**SP**	**SR**
Tester 1	Nokia	Intel	ETRI	Nokia	Intel	Nokia	Nokia	Intel	Nokia
Tester 2	KAU	ETRI	KAU	Intel	ETRI	KAU	KAU	ETRI	Intel

Results of the cross-check:

- The results from KAU were not available for cross-check; hence, the sequences SA, SD, SN and SO have not been cross-checked.
- The cross-checks for SE, and SR have differences

The cross-checks for SB, SJ and SP are perfect.

Results:

Tester 1

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	
ClassroomVid	SA	-40.1%	-37.1%	1.52	-14.9%	-16.0%	-33.0%	-33.5%	Nokia
Museum	SB	-23.7%	-24.1%	16.33	-25.1%	-25.2%	-18.6%	-20.2%	Intel
Fan	SO	-39.8%	-37.3%	7.82	-40.6%	-39.5%	-37.9%	-35.6%	Nokia
Kitchen	SJ	-17.1%	-23.0%	16.40	-18.5%	-24.7%	-24.2%	-24.8%	Intel
Painter	SD	-23.0%	-22.1%	7.86	-22.4%	-23.2%	-19.8%	-19.1%	ETRI
Frog	SE	-30.4%	-30.0%	6.46	-30.4%	-31.0%	-32.5%	-31.4%	Nokia
Carpark	SP	-25.3%	-26.3%	7.38	-25.8%	-27.1%	-20.1%	-23.2%	Intel
Chess	SN	-30.9%	-30.2%	14.75	-31.2%	-30.2%	-32.1%	-31.0%	Nokia
Group	SR	-28.4%	-28.4%	12.15	-28.2%	-28.8%	-24.0%	-24.0%	Nokia

Tester 2

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	
ClassroomVid	SA								KAU
Museum	SB	-23.7%	-24.1%	16.33	-25.1%	-25.2%	-18.6%	-20.2%	ETRI
Fan	SO								KAU
Kitchen	SJ	-17.1%	-23.0%	16.40	-18.5%	-24.7%	-24.2%	-24.8%	ETRI
Painter	SD								KAU
Frog	SE	-30.0%	-29.5%	6.46	-30.0%	-30.5%	-32.1%	-30.9%	Intel
Carpark	SP	-25.3%	-26.3%	7.38	-25.8%	-27.1%	-20.1%	-23.2%	ETRI
Chess	SN								KAU
Group	SR	-28.2%	-28.1%	12.15	-28.2%	-28.7%	-23.9%	-23.8%	Intel

Summary of recommendations from experimenters:

Intel:

- It will be nice to have these steps get integrated in the TMIV software itself to simplify the process and also have the metadata actually being adjusted to reflect the packing information and the presence of packed video data sub-bitstreams while having a single bitstream multiplexing all other components in.

ETRI:

- With respect to A17, significant gains were observed.
- The results from EE 1.a couldn't be used as an anchor, since the configuration file for VVC encoding is different
- More concise and precise experimentation will be possible if different procedures for frame packing (padding, merging, splitting, etc.) are merged into TMIV 7.0.

Exploration Experiment EE2: decoder-side depth estimation

Description: EE2a and EE2b experiments are proposed to investigate the effect of sending 'feature information' determined at the encoder-side and transmitted to the decoder, in order to assist the decoder-side depth estimation algorithm. The anchor is G17, the decoder side depth estimation approach.

Both EEs evaluate the transmission of features that aim at reducing the complexity of the depth estimation on the client side.
The features that were being transmitted for this EE were the following:

- A skip flag, transmitted from a threshold that was the same for all sequences.
- A partition flag, transmitted from a threshold that was the same for all sequences.
- A quantization value for the signalization of Zmin and Zmax, that was ‘256’ for all 16-bit sequences, and ‘4’ for SJ (kitchen, which has 10 bit depths). This value effectively brought all depths down to 8 bits and so can be thought of as sequence independent.

For EE2.a, all features are considered. For EE2.b, only the skip flag feature is considered.

A modified IVDE (available in EE2 branch in MPEG GitLab) makes use of this information. Modifications will be described in [m55751].

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 well matching, besides some small differences in SR in synthesized views (SA showed -0.3% in PSNR but bitrates and quality are the same).

Results for EE2a:

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%	100%	1.07	30	100.0%	100.0%	#DIV/0!	19.9%
Museum	SB	---	---	13.29	-48.1%	-29.3%	-59.1%	-33.3%	100%	1.07	30	100.0%	100.0%	#DIV/0!	19.0%
Fan	SO	-0.4%	3.2%	10.13	4.0%	7.1%	-0.0%	4.1%	100%	1.07	30	100.0%	100.0%	#DIV/0!	67.3%
Kitchen	SJ	-28.4%	-20.4%	12.82	-15.3%	-9.2%	-24.5%	-17.2%	100%	1.07	30	100.0%	100.0%	#DIV/0!	68.8%
Painter	SD	-8.9%	21.3%	8.18	72.5%	71.4%	-21.3%	10.3%	100%	1.07	30	100.1%	100.0%	#DIV/0!	68.6%
Frog	SE	3.5%	13.7%	6.58	10.0%	19.2%	-1.0%	12.6%	100%	1.07	30	100.1%	100.0%	#DIV/0!	27.8%
Carpark	SP	-24.0%	-10.5%	9.59	-15.5%	-3.9%	13.1%	12.7%	83%	0.89	25	100.0%	100.0%	#DIV/0!	81.1%
Chess	SN	######	######	0.00	######	######	######	######	50%	0.53	30	152.4%	######	#DIV/0!	#DIV/0!
Group	SR	---	---	17.93	-62.2%	-39.4%	---	---	100%	1.07	30	100.0%	100.0%	#DIV/0!	48.5%
MIV		---	---	**0.00**	---	---	---	---	93%	0.99		105.9%	######	#DIV/0!	#DIV/0!

Optional content - Proposal vs. Low/High-bitrate Anchors															
Fencing	SL	-35.8%	-15.1%	12.68	-3.0%	6.1%	-3.6%	1.2%	83%	0.89	25	100.0%	100.0%	#DIV/0!	39.5%
Street	SU	14.2%	14.3%	8.12	4.0%	8.9%	8.0%	10.3%	83%	0.89	25	100.0%	100.0%	#DIV/0!	45.3%
Hall	ST	---	---	17.25	133.0%	59.4%	---	---	83%	0.89	25	100.0%	100.0%	#DIV/0!	53.7%
ChessPieces	SQ	######	######	0.00	######	######	######	######	50%	0.53	30	140.4%	######	#DIV/0!	#DIV/0!
Hijack	SC	---	---	20.96	---	-57.7%	---	---	100%	1.07	30	100.0%	100.0%	#DIV/0!	37.0%
MIV		---	---	**0.00**	---	#####	---	---	80%	0.86		108.1%	######	#DIV/0!	#DIV/0!

Results for EE2b:

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	92.9%	27.1%	5.68	157.3%	24.1%	9.4%	8.3%	100%	1.07	30	100.0%	100.0%	######	22.7%
Museum	SB	-99.9%	-0.5%	14.84	-8.4%	-1.1%	-20.0%	-2.0%	100%	1.07	30	100.0%	100.0%	######	22.0%
Fan	SO	5.2%	3.8%	10.31	3.2%	2.4%	5.3%	4.3%	100%	1.07	30	100.0%	100.0%	######	119.1%
Kitchen	SJ	10.0%	8.7%	13.46	8.1%	5.4%	8.7%	8.0%	100%	1.07	30	100.0%	100.0%	######	96.5%
Painter	SD	-43.8%	-26.6%	7.57	-21.0%	-10.4%	-39.3%	-24.2%	100%	1.07	30	100.1%	100.0%	######	273.0%
Frog	SE	-2.5%	-1.5%	6.62	-0.8%	-0.4%	-3.8%	-2.0%	100%	1.07	30	100.0%	100.0%	######	137.3%
Carpark	SP	-5.8%	-4.3%	9.57	-5.6%	-3.3%	4.6%	1.9%	83%	0.89	25	100.0%	100.0%	######	229.8%
Chess	SN	######	######	0.00	######	######	######	######	50%	0.53	30	159.0%	######	######	#DIV/0!
Group	SR	---	-89.3%	22.41	-62.3%	-39.3%	---	-99.6%	100%	1.07	30	100.0%	100.0%	######	107.9%
MIV		---	#####	**0.00**	#####	#####	---	#####	93%	0.99		106.6%	######	######	#DIV/0!

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

Fencing	SL	17.9%	18.9%	12.75	6.4%	3.9%	10.0%	9.1%	83%	0.89	25	100.0%	100.0%	######	104.4%
Street	SU	-0.5%	-0.6%	7.99	-0.6%	-0.5%	0.6%	0.3%	83%	0.89	25	100.0%	100.0%	######	161.1%
Hall	ST	---	---	18.00	---	82.7%	---	---	83%	0.89	25	100.0%	100.0%	######	164.7%
ChessPieces	SQ	######	######	0.00	######	######	######	######	50%	0.53	30	138.9%	######	######	#DIV/0!
Hijack	SC	48.2%	26.6%	22.81	17.7%	7.7%	31.2%	14.0%	100%	1.07	30	100.0%	100.0%	######	43.9%
MIV		---	---	**0.00**	---	#####	---	---	80%	0.86		107.8%	######	######	#DIV/0!

Comments / recommendations:

- For EE2.a, there is a consistent decoder-side complexity reduction, and generally no overall loss of quality.
- 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.
- Orange recommends future investigation of the several aspects of a feature-driven DSDE.
- PUT recommends not to continue EE2.b, as provided results show that the use of all features provides (on average) better quality of decoder-derived depth maps with additional complexity reduction.

Exploration Experiment EE3: MV-HEVC

EE3.a: V17, one atlas per view

Description: V17 configuration is applied, with MV-HEVC instead of HEVC. MV-HEVC is applied independently on the texture and on the depth views, and each view contains only one atlas.

Cross-check: ETRI and Intel have both computed SN and SR. For SN, there is a perfect match. For SR, there are very minor differences.

Results:

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Tester
ClassroomVideo	SA	5.4%	-2.7%	-32.5%	-16.9%	4.0%	-3.8%	ETRI
Museum	SB	7.8%	7.1%			2.3%	3.3%	Tencent
Fan	SO	2.8%	2.4%	1.5%	0.5%	-1.5%	-1.1%	Intel

Kitchen	SJ	-11.2%	-10.5%	-9.0%	-7.4%	-15.9%	-16.8%	Intel
Painter	SD	-18.7%	-18.0%	-25.4%	-24.8%	-23.0%	-23.8%	ETRI
Frog	SE	-20.9%	-18.4%			-23.9%	-22.1%	Tencent
Carpark	SP	-16.9%	-19.6%	-18.7%	-18.7%	-20.5%	-24.7%	Intel
Chess	SN	7.1%	6.2%	27.3%	12.0%	6.2%	3.8%	Intel / ETRI
Group	SR	6.3%	5.6%	1.5%	3.9%	4.3%	2.6%	Intel / ETRI

EE3.b: V17, HEVC replaced by MV-HEVC

Description: V17 configuration is applied, with MV-HEVC instead of HEVC. MV-HEVC is applied independently on the texture and on the depth views, and each view contains several atlases.

Cross-check: ETRI and Intel have both computed SB and SP. For SB, there is a perfect match. For SP, there are differences resulting from a difference of encoding of texture atlas at QP5 only.

Results:

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Tester
ClassroomVideo	SA	6.0%	-2.5%	-32.8%	-16.7%	3.7%	-3.9%	ETRI
Museum	SB	10.2%	9.5%	10.8%	12.2%	3.7%	6.1%	ETRI
Fan	SO	5.9%	5.3%	4.3%	3.4%	1.1%	1.5%	Intel
Kitchen	SJ	11.3%	11.1%	15.9%	14.7%	4.7%	2.3%	Intel
Painter	SD	-8.2%	-8.6%	-8.8%	-9.8%	-9.9%	-11.3%	ETRI
Frog	SE	3.3%	4.7%	1.7%	3.7%	0.8%	2.5%	Intel
Carpark	SP	3.8%	2.1%	2.8%	1.4%	3.1%	-0.9%	ETRI
Chess	SN	-4.4%	1.8%	14.1%	8.6%	3.5%	0.4%	ETRI
Group	SR	5.0%	4.9%	0.8%	2.8%	3.0%	1.9%	Intel

EE3.c: A17, HEVC replaced by MV-HEVC

Description: A17 configuration is applied, with MV-HEVC instead of HEVC. MV-HEVC is applied independently on the texture and on the depth views, and each view contains several atlases.

Cross-check: ETRI and Intel have a perfect match on SR, Tencent and Intel have a perfect match on SE.

Results:

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Tester
ClassroomVideo	SA	6.79%	1.0%	-30.5%	-12.0%	1.9%	2.5%	ETRI
Museum	SB	1.8%	1.4%			-1.02%	-2.1%	Tencent

Fan	SO	0.8%	-0.2%	-0.6%	-1.8%	-1.5%	-2.1%	Intel
Kitchen	SJ	6.1%	4.9%	10.3%	8.9%	4.9%	2.5%	Intel
Painter	SD	-2.2%	-2.3%	-3.9%	-4.3%	-3.3%	-3.5%	ETRI
Frog	SE	-3.3%	-3.2%			-4.1%	-4.6%	Tencent
Carpark	SP	3.3%	3.6%	-4.3%	-4.5%	-4.1%	-4.6%	Intel
Chess	SN	4.6%	1.3%	5.4%	3.2%	0.7%	-3.7%	ETRI
Group	SR	9.3%	4.6%	6.7%	2.7%	3.2%	0.2%	ETRI

Comments / recommendations:

- EE3a can achieve some gains for some contents for some metrics, but degradation is also observed. MV-HEVC outperforms V17 (and A17) for natural perspective content when each view is in its own atlas.
- EE3b is generally worse than EE3a: when multi-view prediction is used, it is better to have only one atlas per view.
- Atlases with patches are not friendly to MV-HEVC (only saw improvement in SE).
- Intel recommends only doing EE3a on mandatory content going forward (but compare with both A17 & V17) to keep an eye on how MIV is improving vs MV-HEVC coding.
- Tencent recommends watching some pose traces on EE3a.

Exploration Experiment EE4: x265 evaluation

Description: This exploration experiment proposes using x265 instead of HEVC.

Cross-check: Tencent did not manage to match PUT's results. Nokia could not perform this experiment.

Results provided by PUT:

Sequence		Mandatory content - Proposal vs. Low/High-bitrate Anchors										Runtime ratio (%)			
		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															
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 / recommendations:

- Replacing HM16.16 by x.265 reduces the video encoding step by a factor of 500, which allows reducing the time needed for entire TMIV encoding 10 times.
- Replacing HM16.16 by x.265 brings huge losses in coding efficiency, except for Museum sequence.
- PUT recommends to re-discuss the possibility of substituting HM with x265 in MIV-related experiments.

Exploration Experiment EE5: IVDE anchor depth generation

Description: The aim of this experiment was to generate a MIV anchor based on the depths maps generated by IVDE.

Cross-check:

- Philips received all depth maps from PUT and will fully crosschecked A17 anchor results. Some mismatches were found, but are believed to relate to undefined behavior in the equirectangular reprojection (TMIV issue #148). After rerunning the TMIV encoder, but compiled with VC16, it was possible to produce matching results for that stage.
- ETRI did the partial crosscheck of the 17 frames of depths maps that were provided by PUT. ETRI have observed the phenomenon that only the first frame is identical and the remained 16 frames are slightly different in the aspect of both visual quality and md5sums. To verify whether the IVDE functionality is compiler depedent, PUT has provided the new depth map corresponding to the single view (v6) of SE based on IVDE compiled with vc15. Then, ETRI has confirmed that new depth is identical including md5sums.

Seq	SA	SB	SD	SE	SJ	SN	SO	SP	SR
Tester 1	PUT	PUT	PUT	PUT	PUT	PUT	PUT	PUT	PUT
Tester 2	Philips	Philips	Philips	Philips	Philips	Philips	Philips	Philips	Philips
Tester 3			ETRI	ETRI		ETRI	ETRI	ETRI	

Results:

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Tester
ClassroomVideo	SA			4.14			688.0%	PUT
Museum	SB			25.09				PUT
Fan	SO	-68.0%	-67.0%	5.62	-55.5%	-58.6%	-51.7%	PUT
Kitchen	SJ	154.4%	94.2%	15.79	202.3%	101.6%	88.2%	PUT
Painter	SD	67.2%	63.2%	7.57	63.1%	61.8%	79.8%	PUT
Frog	SE	-4.8%	-1.1%	5.33	-2.8%	-0.4%	0.2%	PUT
Carpark	SP	47.3%	60.2%	6.99	39.9%	57.4%	51.1%	PUT
Chess	SN			27.26				PUT
Group	SR			28.92		145.7%		PUT

Comments / recommendations:

- As expected, the quality of depth maps generated in the experiment is lower than for CTC depth maps. The depth maps were 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 by PUT.
- PUT recommends that EE5 should be continued to test the performance of the new TMIV 8.0 and new (possible) IVDE 3.0.
- ETRI pointed out that even though there exist some mismatches due to the compiler version of IVDE, it seems that those amounts are minor to make an impact on the final results. This compiler dependent error is already being checked by the software coordinator of IVDE.