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 m56601
April 2021, 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 [WG04N0055], evaluates in a collaborative process the coding of TMIV generated attribute and geometry atlases with different configurations and implementations of VVC and with HEVC. The EE handles in addition the generation of IVDE anchor depth maps and the study of multiple patches per geometry patches.

Related contributions

The detailed experimental results, cross-check results, comments and recommendations are provided by each participant in input contributions as summarized in the table below.

Affiliation	Contribution
ETRI-Immersive Media	M56559
ETRI-Media Codec	M56611
Nokia	M56612
Philips	M56325
PUT	M56563
Tencent	M56324
ULB	M56625

Exploration Experiment EE1: evaluation of various VVenC configurations

EE1.a: “slower” + EE1.b: “faster”

VVenC includes five pre-set configurations: “slower”, “slow”, “medium”, “fast”, “and” faster”. Two configurations are tested, and compared to A17 anchor: “slower” and “faster”. A subset of sequences are considered: SB, SD, SJ, SN, SO, SP mixing computer generated and natural content, equirectangular and perspective.

Participants

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

Seq	SB	SD	SJ	SN	SO	SP
Tester 1	Tencent	Tencent	Tencent	Tencent	Philips	Philips
Tester 2	Philips	Philips	ETRI	ETRI	ETRI	ETRI

Results of the cross-check:

The cross-check highlights differences for all test sequences. They are supposed to come from the different compilers used or the different number of threads used.

EE1a:

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR		
Museum	B	0.4%	-0.1%	Philips	Tencent
Fan	O	-0.3%	-0.6%	Philips	ETRI
Kitchen	J	-0.1%	0.1%	ETRI	Tencent
Painter	D	-0.9%	-0.8%	Philips	Tencent
Carpark	P	-0.7%	-0.7%	Philips	ETRI
Chess	N	-0.7%	-0.6%	ETRI	Tencent
MIV		**-0.4%**	**-0.4%**		

EE1b:

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR		
Museum	B	0.1%	0.2%	Philips	Tencent
Fan	O	0.5%	0.1%	Philips	ETRI
Kitchen	J	-0.2%	-0.3%	ETRI	Tencent
Painter	D	0.3%	0.2%	Philips	Tencent
Carpark	P	-0.8%	-0.6%	Philips	ETRI
Chess	N	0.1%	0.1%	ETRI	Tencent
MIV		**0.0%**	**-0.0%**		

Results:

The tables below reports the results as produced by the experimenters. Using the “slower” configuration improves the anchor by 5.1% to 5.6% on average, while increasing the encoder runtime by a factor 8. Using the “faster” configuration degrades the anchor significantly, while decreasing the encoder runtime by a factor 3.

EE1a:

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	TMIV encoding	Video encoding
Museum	B	-1.6%	-2.9%	233.4%	411.3%
Fan	O	-7.2%	-6.8%	601.0%	680.2%
Kitchen	J	-4.6%	-6.1%	388.8%	388.8%
Painter	D	-4.0%	-4.0%	479.0%	479.0%
Carpark	P	-5.3%	-5.1%	694.2%	732.4%
Chess	N	-8.0%	-8.8%	2439.0%	######
MIV		**-5.1%**	**-5.6%**	805.9%	855.1%

EE1b:

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	TMIV encoding	Video encoding
Museum	B	80.2%	91.3%	56.3%	17.6%
Fan	O	166.3%	128.3%	27.5%	16.0%
Kitchen	J	57.9%	62.0%	11.1%	11.1%
Painter	D	107.1%	69.4%	25.7%	16.1%
Carpark	P	186.0%	158.1%	21.0%	15.9%
Chess	N	143.4%	140.9%	68.4%	68.4%
MIV		**123.5%**	**108.4%**	35.0%	24.2%

Summary of recommendations from experimenters:

ETRI:

- Minor BD-rate differences were observed at crosschecked results, but these gaps do not affect examining the general behavior of “slower” and “faster” VVenC configuration.
- In case of “slower” configuration, the BD-rate gain is a way too small (maximum 12%) compared to the time increment (minimum 3.8 times to maximum 27.4 times). Therefore, it is not worth performing further experiment.
- In case of “faster” configuration, the range of BD-rate loss is between 47% to 200% and the encoding time is reduced by minimum 2 times to 10 times. If the time reduction is more valuable than BD-rate loss, the group can consider this configuration as one possible candidate.

Tencent:

- Keep the “slow” configuration of VVenC as the current anchor.

Exploration Experiment EE2: TMIV + different 2D codecs and configurations

EE2.a: HM validation + EE2.b: VTM validation

The goal of this experiment is twofold:

- To make sure that the new adoptions and their implementations in the reference software remain compatible with a variety of other 2D codecs, and their configurations; the tested codecs and configurations are different from the one used to generate the CTC anchors.

- Evaluate performance of other codecs and configurations compared to the current VVenC-based CTC anchors.

The anchor for this experiment is the CTC A17 configuration. The sequences SB, SD, SJ, SN, SO, and SP are used. The choice of sequences reflects a good mix of synthetic and natural content, as well as content with equirectangular and perspective projections.

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

Seq	**SB**	**SD**	**SJ**	**SN**	**SO**	**SP**
Tester 1	Tencent	Tencent	Tencent	Tencent	Philips	Philips
Tester 2	Philips	Philips	Nokia	Nokia	Nokia	Nokia

Results of the cross-check:
With the HM, there is a perfect match for all sequences.

EE2a:

Sequence		**High-BR BD rate Y-PSNR**	**Low-BR BD rate Y-PSNR**		
Museum	B	-0.0%	-0.0%	Philips	Tencent
Fan	O	0.0%	0.0%	Philips	Nokia
Kitchen	J	0.0%	0.0%	Nokia	Tencent
Painter	D	0.0%	0.0%	Philips	Tencent
Carpark	P	0.0%	0.0%	Philips	Nokia
Chess	N	-0.1%	-0.0%	Nokia	Tencent
MIV		**-0.0%**	**-0.0%**		

EE2b:

Sequence		**High-BR BD rate Y-PSNR**	**Low-BR BD rate Y-PSNR**		
Museum	B	0.2%	0.0%	Philips	Tencent
Fan	O	-0.0%	-0.0%	Philips	Nokia
Kitchen	J	0.1%	-0.2%	Nokia	Tencent
Painter	D	-2.0%	-1.5%	Philips	Tencent
Carpark	P	-0.0%	0.0%	Philips	Nokia
Chess	N	1.5%	0.3%	Nokia	Tencent
MIV		**-0.0%**	**-0.2%**		

Results:

EE2a:
Using the HM16.16 degrades the anchor by a factor 3, while increasing the encoder runtime by a factor 2.5.

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	TMIV encoding	Video encoding
Museum	B	30.7%	30.2%	120.6%	138.9%
Fan	O	50.5%	46.0%	115.1%	117.5%
Kitchen	J	17.0%	24.6%	147.4%	147.4%
Painter	D	25.5%	24.4%	114.1%	115.9%
Carpark	P	35.6%	33.6%	112.4%	113.2%
Chess	N	36.0%	29.0%	845.6%	845.6%
MIV		**32.6%**	**31.3%**	242.5%	246.4%

EE2b:
Using the VTM improves the anchor by 5.5% to 6.8% on average, while increasing the encoder runtime by a factor 15.

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	TMIV encoding	Video encoding
Museum	B	-2.0%	-4.0%	451.9%	763.9%
Fan	O	-10.1%	-9.7%	825.0%	939.5%
Kitchen	J	-5.0%	-7.4%	819.8%	819.8%
Painter	D	-4.9%	-4.6%	830.4%	924.6%
Carpark	P	-2.6%	-4.1%	874.8%	924.6%
Chess	N	-8.4%	-11.2%	5396.5%	######
MIV		**-5.5%**	**-6.8%**	1533.1%	######

Summary of recommendations from experimenters:

Tencent:

- Keep the “slow” configuration of VVenC as the current anchor.

Exploration Experiment EE3: IVDE anchor depth generation

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

Participants: The workload of conducting this experiment was split, based on the cross-checked stage of the experiment, among the participants in the following way:

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	ETRI	ETRI	ETRI	ETRI

Tester 1 estimated depth maps using IVDE 3.0 and tested them in TMIV 8.0 using MIV anchor configuration. Tester 2 cross-checked the MIV anchor after receiving depth maps from Tester 1, while Tester 3 cross-checked estimated depth maps only.

Results of the cross-check:

The cross-check of the TMIV encoding using provided depth maps is shown below. Some differences can be seen for omnidirectional sequences, which is a known TMIV issue related to floating-point representation.

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
ClassroomVideo	0.0%	0.0%	4.57	-0.0%	-0.0%	0.0%	-0.0%
Museum	-0.1%	-0.0%	24.74	-0.2%	0.0%	-0.0%	-0.0%
Fan	0.0%	0.0%	5.91	-0.0%	-0.0%	0.0%	0.0%
Kitchen	-0.0%	-0.0%	16.06	-0.0%	-0.0%	-0.0%	-0.0%
Painter	-0.0%	-0.0%	7.98	0.0%	0.0%	-0.0%	-0.0%
Frog	0.0%	0.0%	5.63	-0.0%	-0.0%	0.0%	0.0%
Carpark	0.0%	0.0%	7.33	0.0%	-0.0%	0.0%	0.0%
Chess	0.6%	-0.1%	28.52	-0.1%	0.0%	0.1%	0.1%
Group	-0.0%	-0.0%	22.28	-0.0%	0.0%	0.0%	-0.0%
MIV	**0.0%**	**-0.0%**	**13.67**	**-0.0%**	**0.0%**	**0.0%**	**0.0%**

The cross-check of depth maps (from ETRI-Immersive Video report): in case of perspective sequences, the md5sum values were identical across all views, but differences were observed for omnidirectional sequences across most of views (there were some views that were identical). However, only minor visual distinctions were observed. It was inferred that there still exist minor compiler dependent factors when IVDE dealing with omnidirectional inputs.

Results:

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

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
ClassroomVideo	A	---	---	4.57	---	---	699.8%	663.1%
Museum	B	---	---	24.74	---	---	---	---
Fan	O	-66.7%	-65.7%	5.91	-53.2%	-56.7%	-48.4%	-53.2%
Kitchen	J	189.5%	95.8%	16.06	294.3%	106.9%	87.8%	56.3%
Painter	D	52.6%	48.6%	7.98	47.5%	46.3%	63.8%	52.9%
Frog	E	-5.6%	-1.0%	5.63	-2.6%	0.4%	0.7%	2.3%
Carpark	P	44.2%	57.1%	7.33	32.5%	52.5%	46.4%	58.5%
Chess	N	---	---	28.52	---	---	---	---
Group	R	---	---	22.28	233.0%	22.7%	---	---
MIV		**---**	**---**	**13.67**	**---**	**---**	**---**	**---**

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

Fencing	L	-8.8%	19.3%	9.05	35.8%	37.6%	1.7%	23.2%
Hall	T	-59.9%	-49.1%	9.61	-59.4%	-48.0%	-48.1%	-43.3%
Street	U	31.1%	36.5%	8.80	-3.3%	21.3%	29.6%	37.2%
ChessPieces	Q	---	---	28.35	---	---	---	---
Hijack	C	---	---	22.44	---	199.5%	---	---
Mirror	I	---	-48.0%	9.02	---	-38.9%	---	-45.6%
MIV		**---**	**---**	**14.54**	**---**	**---**	**---**	**---**

Summary of comments and recommendations from experimenters (only PUT provided the experiment-related comments)**:**

- Recommendation: EE3 should be continued to test the performance of the new TMIV 9.0.
- 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.
- The high quality in SI is the result of mirrors that are present in the scene. In the ground-truth depth maps, the depth of mirrors shows the distance from the camera to the mirror, while in estimated depth, the distance from the camera to the reflected object.

Exploration Experiment EE4: multiple texture patches per geometry patches

Description:

The goal of this experiment was to evaluate the use of packing multiple texture patches for a given geometry patch in order to better render non-lambertian (specular) surfaces. Rendering of

specular regions require that the texture be adapted to the view-orientation of the observer. Towards this goal two experiments were proposed in m55977.

- Additional atlases for specular patches
- Multiple texture patches for a geometry patch

In the first experiment, the goal was to code addition texture patches along with their geometry data (depth) redundantly. This would give a base-line case for evaluation. The second experiment was to not code the redundant geometry, thus evaluating a projected savings in bitrate. Due to time constraints the second experiment could not be completed. However, the first experiment was completed. Additional experiments, that were based on the first experiment and proposed by the group in an ad-hoc call, was also performed.

The proposer of this experiment was Nokia, and they were requested to provide a software for experimenting which was provided in m56338.

Additional to a cross-check of m56338, ETRI-Immersive media also report on two other experiments they have conducted in m56559. These experiments were

- performance comparison between [m56338] and TMIV 8.1 with setting the value of parameter "maxLumaError" as half and
- performance comparison between [m56338] and TMIV 8.1 with setting the values of "maxLumaError" and "maxColorError" as half. The main aim of sub-experiment

These experiments were done to investigate if controlling the pruning relevant parameters in TMIV can handle the specular regions as efficient as m56338.

Cross-check:

. The other participants of the experiment where:

- ETRI-Media Codec
- ETRI-Immersive Media
- ULB

Rephrased report from ETRI-Immersive Media (m56559)

- Ignorable BD-rate differences were observed during crosscheck.
- Under the same amount of un-pruned luma samples per frame, it seems like m56338 can handle specular regions more efficiently than the current TMIV structure
- It is worth further pursuing this experiment and it will be interesting to watch pose traces to examine how much improvement can be achieved at specular regions

Report from ETRI-Media-Codec (m56611)

- It was informed that there were differences during cross-check. Furthermore, they also report that in their experimentation of using m56338 with non-specular content (seq A, and O) m56338, performed better in higher bitrate and worse in lower bitrate. They further report that subjectively annoying artifacts was found in sequence A, while they found improvements in sequence O. They suggest a continuation of study for this topic.

Report from ULB (m56625)

- ULB have reported cross-checks on Mirror (SI), Fan (SO) and ChessPieces (SQ). They have also informed that they are in the process simulating Kitchen (SJ) and will provide their findings soon. They report, generally, their results are consistent with the results described in m56338. They had a concern about an increase in processing time with respect to the Anchor. For SQ, ULB has obtained different results in quality metrics of virtual views, with an average of -10dB discrepancy compared to the results of m56338 which is intriguing because the intermediate results in the CTC pipelines match. They recommend further study on this topic.

Results:

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
ClassroomVideo	A	-7,8%	8,9%	1,00	-6,1%	10,2%	15,9%	21,8%
Fan	O	-6,5%	-0,2%	6,25	3,9%	6,7%	0,8%	5,6%
Kitchen	J	-52,1%	-19,8%	13,03	-50,7%	-12,7%	-44,5%	-17,2%
ChessPieces	Q	---	-62,0%	13,50	-73,5%	-22,5%	---	-67,0%
Mirror	I	-61,0%	-27,5%	6,67	-22,3%	-4,9%	-40,0%	-18,5%

Comments / recommendations:

- At least objectively, the experiment performed shows gains. The proponents of this experiment propose to continue with this experiment, focusing on further improvements by not coding the redundant geometry at all.
- At least one of the non-proponent experimenters proposed to continue with this experiment.