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 57276
July 2021, Online

Title: **Outcomes of exploration experiments – summary report**
Source: **Vinod Kumar Malamal Vadakital (Nokia),**
Joel Jung (Tencent),
Dawid Mieloch (PUT)

Abstract
This contribution is a summary of outcomes of all experiments listed in N0088. A total of 10 organizations participated in one or more of the listed experiments. Three main experiments were agreed upon, of which two experiment had sub-experiments. Significant participation and engagement from experimenters were observed, and several useful recommendations are provided from participating organizations.

1 Introduction

Three main exploration experiments, with two of the three experiments having sub-experiments, were agreed upon in MPEG-134. These experiments are listed below:

- [EE-1]: MIV anchor generation using depth maps using IVDE
- [EE-2]: Investigation for anchor generations for verifications tests
  - [EE-2.1]: Investigate the coding of MIV scenes using MV-HEVC
  - [EE-2.2]: Investigate the coding of MIV scenes using 3D-HEVC
- [EE-3]: Investigation of coding specular content in MIV
  - [EE-3.1]: validation of pruning specular regions
  - [EE-3.2]: validation of packing patches with without redundant geometry

The following member organizations agreed to be part in one or more of the conducted experiments: ETRI-Immersive Media, ETRI-Media Codec, Nokia, Orange, Philips, PUT, Tencent, ULB, ZJU, Interdigital

Additionally, Intel helped in providing the configuration files and instructions for properly conducting the experiment EE-2.1. Gitlab issues and emails were the primary modes of communication between experimenters.

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

m57069	Philips results for Exploration Experiments on Future MPEG Immersive Video
m57087	Tencent results for Exploration Experiments on Coding for Future MPEG Immersive Video
m57168	InterdigitalResultsToExplorationExperiment_EE2_1
m57169	Exploration Experiments on Future MIV: PUT results
m57264	Outcomes of exploration experiments conducted by Nokia
m57130	ETRI results for Exploration Experiments on Future MIV

2 Outcomes of exploration experiments

2.1 EE1: IVDE anchor depth generation

This experiment generates a MIV anchor based on depth maps obtained with IVDE 4.

2.1.1 Participants

- PUT (providing depths for all sequences and applying TMIV)
- Tencent (providing depths for all sequences and applying TMIV)
- Philips (cross check of the TMIV part)
- ETRI + Orange (cross check of depth maps provided by PUT)

2.1.2 Cross-check

- Depth maps: Tencent and PUT have generated all depths maps. There is a perfect match except for sequence SA, for which slight visual differences can be observed, while the configuration files were verified and correct. The reason for this mismatch is probably floating-point operations.
- TMIV results: Philips, PUT and Tencent have generated the TMIV results from IVDE4 depth maps. The results are perfectly matching PUT’s and Philips’ results, except for sequence SA. The difference in the depth computation yield to a 2.2% BD-rate difference at high bitrates, and 1.0% difference at low bitrate.
- ETRI had estimated 17 frames' depth maps from sequence D, E, P, N, R based on IVDE and compared them with PUT's in the aspect of md5sums. In case of perspective sequences (D, E, P, and R), the crosscheck was successful since md5sum values were identical across all views. In case of ERP sequence (N), the md5sum results were mostly different across views, but only minor visual differences were observed. As it has already been known within group, it is highly likely caused by compiler difference

2.1.3 Results

According to objective results (Table 2), EE1 seems better than the anchor A17 configuration for SO, SE and ST.

Table 2: Objective results of EE1 in comparison with the MIV anchor (A17 configuration)

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 IV-PSNR**	**Low-BR BD rate IV-PSNR**
ClassroomVideo	A	---	637.0%	4.35	494.1%	391.8%
Museum	B	---	---	18.30	1151.6%	547.2%
Fan	O	-63.1%	-62.0%	5.98	-40.3%	-44.7%
Kitchen	J	80.3%	52.7%	14.36	49.4%	36.2%
Painter	D	43.5%	35.4%	9.07	53.2%	38.7%
Frog	E	-14.7%	-6.1%	6.23	-4.3%	0.4%
Carpark	P	34.8%	45.5%	6.95	40.1%	48.9%
Chess	N	---	---	27.13	---	---
Group	R	---	---	23.38	---	---
MIV		**---**	**---**	**12.86**	**---**	**---**
Optional content - Proposal vs. Low/High-bitrate Anchors						
Fencing	L	1.4%	22.8%	8.95	2.9%	24.9%
Hall	T	-63.8%	-54.6%	10.55	-45.7%	-45.9%
Street	U	13.4%	18.4%	8.34	16.5%	21.0%
ChessPieces	Q	---	---	27.71	---	---
Hijack	C	---	---	22.21	---	---
Mirror	I	0.0%	0.0%	8.77	0.0%	0.0%
MIV		**---**	**---**	**14.42**	**---**	**---**

PUT had run the experiment again for these sequences in the A97 configuration to evaluate if new depth maps can be provided for CTC. SO is already using IVDE-generated depth maps, so the outcome was the same as in the A97 anchor. For SE and ST no meaningful difference between pose traces were observed after initial viewing performed by PUT.

2.1.4 Recommendations from experimenters

Philips

- Recommends continuing EE but for relevant sequences only.
- In the next iteration of the EE, use 17 frames for coding and 97 frames for rendering pose traces

PUT

- Recommends continuing to test the performance of the new TMIV 10.0 if considerable changes will be made.

Tencent

- Recommends doing a viewing on the 97 frames pose-traces for SO, SE, ST and for the group to decide whether there is sufficient evidence to replace the current depth maps with the IVDE4 depth maps.

2.2 EE2.1: MV-HEVC

With a view of producing anchors for the verification tests, the goal of this experiment was to identify the appropriate simulation pipeline and have an initial performance evaluation of using the Multi-View High Efficiency Video Codec (MV-HEVC). All mandatory sequences described in the CTC were considered for this experiment. The allocation of sequence to each organization is shown in Table **3**.

Table 3: Sequence allocation

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

For this experiment, there was a need to have a method to choose the sub-set of camera sequences that is used for coding. This method was provided, implemented as a python script, and the details of which can be found in m57068.

Next, the software packages, their configuration parameters and the process for simulation had to be established. Intel provided the configuration parameter for MV-HEVC and the process for simulation. Philips and ULB provided the configuration parameters for the use of RVS for rendering in decoded source bitstreams.

2.2.1 Cross-check

- ETRI cross-checked the results of sequence E, N and R provided by Interdigital and Nokia. In case of E and N, QPs for MIV View Anchor were identically used. In case of R, tester and cross-checker intentionally used different QPs for choosing the most appropriate QPs for the next verification test. About E, the results including bitrates and metric values were perfectly aligned
- About N, the bitrates were matched, but very minor deltas (mostly 0.00X) were occurred at metric values. It was highly likely caused by compiler-dependent behavior of RVS on ERP input. About R, RD-curves were strongly overlapped between tester and cross-checker. Therefore, both results seem to be generated in the right way

2.2.2 Results

According to objective results (Table 4), MV-HEVC provides worse quality than the MIV anchor in A17 configuration for all tested sequences.

Table 4: Objective results of EE2.1 in comparison with the MIV anchor (A17 configuration, green cells indicate that MIV performs better)

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 IV-PSNR	Low-BR BD rate IV-PSNR
ClassroomVideo	A	-89.1%	-47.7%	0.99	-65.4%	-37.4%
Museum	B	-47.9%	-28.3%	9.50	-44.7%	-23.1%
Fan	O	---	-58.9%	8.04	-50.7%	-41.9%
Kitchen	J	---	---	14.81	---	---
Painter	D	-78.0%	-72.8%	8.14	-79.8%	-74.1%
Frog	E	---	-60.7%	7.38	---	-58.4%
Carpark	P	---	-84.2%	7.10	-89.9%	-75.8%
Chess	N	---	---	13.16	---	---
Group	R	-63.6%	-42.2%	13.03	-71.7%	-41.4%
MIV		**---**	**---**	**9.13**	**---**	**---**

2.2.3 Recommendations from experimenters

Philips:

- Recommends continuing EE2.1 using 97-frame bitstreams and 300-frame pose traces, and updated QP's.
- Recommends a renewed call for test material with non-planar camera rigs.
- Recommends creating a Sequence Configuration 3.0 format based on the CTC configuration files and add support in TMIV 10 and RVS 4.
- Recommends improve the description of all exploration experiments. For instance, it was not clear if and how pose traces for EE1 and EE2.1 had to be generated. A GitLab issue AHG#312 has be opened to promote discussion on this topic.

Interdigital:

- Recommends using the results provided in their contribution as basis for future comparison with MIV.

PUT:

- Recommends continuing to explore the performance of MV-HEVC in comparison with TMIV.

Nokia:

- Recommends fine-tuning this experiment to generate good verification anchors

2.3 EE2.2: 3D-HEVC

Since there was no expertise provided by member organization to progress with this experiment, this experiment was not conducted.

2.3.1 Recommendations from experimenters

Philips:

- Recommends dropping of this experiment, as they have no expertise on this topic and cannot proceed without inputs from experts in 3D-HEVC.

2.4 EE3.1 Multiple texture patches per geometry patches - Pruning validation

Non-Lambertian surfaces are typical in most natural and computer graphics generated scenes. For non-Lambertian surfaces the appearance varies with the observer's viewpoint. Hence, the aim of the experiment was to explore the use of multiple texture patches per geometry patch to synthesize view-dependent effects of non-Lambertian surfaces. The hypothesis was that identifying and segmenting regions exhibiting non-Lambertian characteristics required adaptation of the pruner module in TMIV. The adapted pruner was provided by Nokia and the experiment was to validate that the adapted pruner would behave properly also in scenes where non-Lambertian surfaces are not prevalent.

The allocation of sequences that were agreed upon in MPEG-134 is shown in Table **5**

Table 5: Sequence allocation

Content Type	**Synthetic Content non-Lambertian**			**Synthetic Content Lambertian**		**Natural Content**		
Seq	**SI**	**SJ**	**SQ**	**SA**	**SO**	**SD**	**SE**	**SP**
Tester 1	ETRI-MC	ETRI-MC	ULB	ETRI-MC	ULB	Nokia	ZJU	Nokia
Tester 2	Nokia	ETRI-IM	ETRI-IM	ZJU	Nokia	ETRI-IM	Tencent	Tencent

2.4.1 Cross-check

At the time of writing this document, only the cross-check of Tencent was available. Tencent's results was a perfect match with Nokia's result. ETRI-IM has conducted the experiment, but a cross-check is not available.

2.4.2 Results

Table 6: Partial objective results of EE-3.1 in comparison to TMIV-9.0 (3 atlases used)

1.070 GP/s

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	Max delta 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
Mandatory content - Proposal vs. Low/High-bitrate Anchors										**Runtime ratio (%)**		
ClassroomVideo	A	######	######	0.00	######	######	0 %	0.00	30	######	######	#DIV/0!
Museum	B	######	######	0.00	######	######	63 %	0.67	30	######	######	#DIV/0!
Fan	O	-16.6%	-8.6%	6.22	-3.5%	1.0%	87 %	0.94	30	######	######	#DIV/0!
Kitchen	J	######	######	0.00	######	######	62 %	0.67	30	######	######	#DIV/0!
Painter	D	2.9%	10.1%	8.68	10.9%	15.7%	88 %	0.94	30	######	######	#DIV/0!
Frog	E	######	######	0.00	######	######	56 %	0.60	30	######	######	#DIV/0!
Carpark	P	-23.3%	-3.4%	4.57	-4.2%	9.4%	78 %	0.84	25	######	######	#DIV/0!
Chess	N	######	######	0.00	######	######	63 %	0.67	30	######	######	#DIV/0!
Group	R	######	######	0.00	######	######	0 %	0.00	30	######	######	#DIV/0!
MIV		**#####**	**#####**	**0.00**	**#####**	**#####**	55 %	0.59		######	######	#DIV/0!
Optional content - Proposal vs. Low/High-bitrate Anchors												
Fencing	L	######	######	0.00	######	######	0 %	0.00	25	######	######	#DIV/0!
Hall	T	######	######	0.00	######	######	0 %	0.00	25	######	######	#DIV/0!
Street	U	######	######	0.00	######	######	0 %	0.00	25	######	######	#DIV/0!
ChessPieces	Q	######	######	0.00	######	######	0 %	0.00	30	######	######	#DIV/0!
Hijack	C	######	######	0.00	######	######	0 %	0.00	30	######	######	#DIV/0!
Mirror	I	-11.4%	0.9%	7.53	3.0%	9.7%	0 %	0.00	30	######	######	#DIV/0!
Trio	V	######	######	0.00	######	######	0 %	0.00	30	######	######	#DIV/0!
MIV		**#####**	**#####**	**0.00**	**#####**	**#####**	0 %	0.00		######	######	#DIV/0!

2.4.3 Recommendations from experimenters

Nokia:

- Recommends combining the two sub-experiments (3.1 and 3.2) into a single experiment that continues investigating efficient compression of non-Lambertian scenes.

Tencent:

- Recommends continuing study of this topic in an EE.

ETRI-IM:

- Recommends continuing study of this topic in an EE.

2.5 EE3.2 Multiple texture patches per geometry patches - Packing validation

In EE-3.1 only the pruner was adapted. In other words, in EE-3.1, when non-Lambertian regions of the scene were identified, the pruner took this into account and multiple textures were coded along with their geometry were coded. This would result in redundancy in geometry patches. This experiment was based on the hypothesis that if the multiple textures had the same geometry, they need not be coded by simply referenced, thus leading to better usage of bandwidth.

Additionally, ULB had produced content with non-Lambertian surfaces in m56450. ULB had agreed to provide their content in a TMIV compatible format.

Table 7: sequence allocation

Seq	Single depth map per view				Multiple depth maps per view	
	SI	SJ	SP	SQ	ULB (m56450 Transparent)	ULB (m56450 Mirror)
Tester 1	ETRI-MC	ETRI-MC	Tencent	ETRI-IM	ULB	ULB
Tester 2	Nokia	ETRI-IM	Nokia	Tencent	ETRI-IM	Nokia

2.5.1 Cross-check

Due to the late release of the software as well as some problems faced by Nokia with producing the right configuration for experiments in this EE, the cross-checks have been delayed or mismatched configuration files have been used.

2.5.2 Results

Table 8: Partial objective results of EE-3.2 in comparison to TMIV-9.0 (3 atlases used)

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 IV-PSNR	Low-BR BD rate IV-PSNR	Pixel rate [%]	Pixel rate [GP/s]	Frame rate [Hz]
Carpark	P	63.2%	100.6%	7.47	46.3%	90.8%	52%	0.56	25

Optional 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 IV-PSNR	Low-BR BD rate IV-PSNR	Pixel rate [%]	Pixel rate [GP/s]	Frame rate [Hz]
ChessPiec	Q	45.5%	-0.4%	19.44	---	75.9%	88%	0.94	30
Mirror	I	43.2%	23.2%	10.24	36.6%	18.7%	0%	0.00	30

2.5.3 Recommendations from experimenters

Nokia:

- Recommends combining the two sub-experiments (3.1 and 3.2) into a single experiment that continues investigating efficient compression of non-Lambertian scenes.

ETRI-IM:

- Recommends continuing study of this topic in an EE.