

**INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 4
MPEG VIDEO CODING**

ISO/IEC JTC 1/SC 29/WG 4 m62472
April 2023, Antalya

Title: AHG on MPEG immersive video (MIV)
Source: Bart Kroon (Philips), Dawid Mieloch (PUT), Joel Jung (Tencent)

1 Abstract

This document is the report of the AHG on MPEG immersive video (MIV).

2 Introduction

- Output documents:
 - [Open output documents by due date](#)
 - [All output documents by meeting](#)
- Mandates:
 - Prepare WD5 of ISO/IEC 23090-12 MPEG immersive video Ed. 2
 - Prepare verification tests for MPEG immersive video
 - Define common test conditions for MPEG immersive video
 - Carry out exploration experiments relating to MPEG immersive video
 - Improve rendering quality of MPEG immersive video
 - Contribute to the integration of MPEG immersive video in MPEG scene description
 - Promote MPEG immersive video outside of MPEG
 - Study new technologies for immersive video coding
- Chairs:
 - [@bartkroon](#) (chair)
 - [@dmieloch](#) (vice-chair)
 - [@joelb.jung](#) (vice-chair)
- Reflector: mpeg-i-visual@lists.aau.at
- Subscribe: <https://lists.aau.at/mailman/listinfo/mpeg-i-visual>
- Offline discussion: [MPEG/MIV/InputDocuments](#)
- Meetings (in [MPEG calendar](#), Video tab):
 - MIV + AG5: February 7 (VT plan), March 21 (progress on renderer and Balloons)
 - MIV + SD: February 13 (14:30 - 16:00 UTC) @ SD link
 - MIV: February 21, March 7, April 4, 18, 19
 - MIV F2F meeting: Sunday April 23, Antalya
 - All calls are **15:00 – 16:30 UTC** unless specified otherwise

- Zoom: [Join from PC, Mac, Linux, iOS or Android](#), password: MPEGIVisua

This is an AHG of ISO/IEC JTC 1/SC 29/WG 04 *MPEG Video coding*, and the Convenor is the convenor of WG 04.

Active AHG guidelines are: [AG 02 N 0046](#)

GitLab registration process:

1. Register an account at <https://mpeg.expert/software> – pending approval
2. Inform the Convenor to request the account approval and join the MIV group as [Reporter or Developer](#)

To register an input document that fits with the mandates of this AHG:

- Register for the upcoming WG 04 meeting as a delegate of your national body
- Register and upload to the [MPEG document management system \(MDMS\)](#)
 - Prefix the title with "[MIV]"
 - Set standard to MPEG-I.
 - Set working group to the best fitting WG/AG → when not WG 04, the AHG chairs may consult with the Convenor if discussion is appropriate.
- Register on GitLab using [this web form](#)
 - The proponent is welcome to open the GitLab issue early to solicit feedback or informally provide initial results.

We welcome non-members to get involved:

- Please request permission of the Convenor to participate.
- The information at <https://mpeg-miv.org> should get you up to speed.

2.1 Related repositories

- Specifications:
 - V3C + V-PCC: [Visual volumetric video-based coding and video-based point cloud compression 2nd edition](#) (WG 07)
 - MIV: [MPEG immersive video](#)
 - Conformance: [MIV conformance and reference software](#)
- Reference software:
 - MPEG-internal: [TMIV](#), [IV-PSNR](#), [IVDE](#), [RVS](#)
 - Public mirrors: [TMIV](#), [IV-PSNR](#), [IVDE](#), [RVS](#)
- MPEG Systems (WG 03):
 - [Systems technologies for volumetric media](#)
 - [MPEG-I Scene description](#)

2.2 Way of working

This AHG has a way of working as described in the following subsections.

2.2.1 Design principles

- The AHG/BoG's use only GitLab (this project).
- Short AHG/BoG reports are crafted to present recommendations to the convenor.

2.2.2 Issue templates

The way of working is to a large extent encoded in issue description templates. These templates include instructions and checklists.

Click on any of the links to open a new issue using the respective template:

Template	Creator	Editor	Provide comments
AHG mandate	AHG chair	AHG chair	AHG member
Input document	Proponent	Proponent, AHG chair	AHG member
Output document	BoG chair	Editor, BoG chair	MPEG expert
Exploration experiment	EE owner	EE owner	MPEG expert

2.2.3 GitLab labels

The definition of a label can be read by hovering over the label. There are three types of labels:

Label type	Responsible	Purpose
Signal progress	Depends on the label	Communication between proponents and chairs, track progress in meetings
Signal recommendations	AHG or BoG chair	Record recommendations for presentation in the Video plenary
Signal decisions	BoG chair	Communication between the BoG chairs and the convenor

To signal decisions:

- The AHG/BoG chairs set **ProbablyAgreement** when consensus has been reached (and there is at least one recommendation)
- The recommendation is presented in the WG 04 plenary

- When accepted, the label **ProbableAgreement** is removed and **Accepted** is added
- The other MPEG-wide "Chair use only" labels are currently not used by the WG 04 convenor

The entire activity can be viewed as a finite state machine. The real-time status is visible on [the board](#).

2.2.4 Meeting notes

The AHG or BoG chair captures notes on a mandate or document as a summary comment on the GitLab issue. To make the comment stand out, the first line starts with a single # for Heading 1 formatting, followed by a title such as "Notes of the AHG call", "Notes of the BoG call" or "Summary".

For clarity, the most recent comment has to capture *all* recommendations (not just the delta):

1. In the comment thread starting from the bottom, search for the latest summary note.
2. Copy the contents of that note.
3. Start a new comment and paste in the old notes.
4. Edit the notes as the discussion progresses.
5. Add/remove relevant labels (use /label and /unlabel for transparency).
6. In a call, show the preview and ask for objection, if none, submit the comment.

The purpose is that:

- It is clear to all involved what the consensus position is.
- The report can be made quickly because only the final summary notes need to be copied (no editing work).
- With multiple summary notes the chronological consensus formation process is visible.

2.3 Previous meeting reports

- [MPEG 131 and newer](#)
- [MPEG 130 Alpbach](#)
- [MPEG 129 Brussels](#)
- [MPEG 128 Geneva](#)
- [MPEG 127 Gothenburg](#)
- MPEG 126: [\(1\)](#) [\(2\)](#) [\(3\)](#)
- [MPEG 125 Marrakech](#)

3 Meeting notes

3.1 MIV verification test preparations

3.1.1 April 4

- An experiment was conducted to improve the best reference using some post filtering of the viewport
- The reported differences were small and it is judged that it would not be useful to repeat the verification test run just changing this aspect
- There is a concern that some pose trace videos were rated lowly because the source material itself was already of a low quality (e.g. blurry texture)
- The MIV verification test may show that VQA of immersive/volumetric media in general may need to be performed differently... for discussion
- There is a discussion on the goals of the verification test:
 - Verify that MIV improves over the anchor based on existing technology (MV-HEVC); this is the official goal -> successfully shown
 - Reflect on the performance of the technology in a meaningful application scenario -> the MOS scores including the best reference are on the low side
- The test was executed correctly. We agree that we are limited by the available source material
- It may be beneficial for MIV 2 to have assessment involving real-time rendering

The AHG recommends:

- Write the final verification test report (WG4 output document) based on the information provided by AG5:
 - Explain how the test was done
 - Discuss the results
- Earlier is better; a draft needs to be discussed with AG5 during the MPEG week.
- Discuss in a MIV 2 BoG how to conduct visual quality assessment of immersive media; involve or report to AG5 during the MPEG week.
- Watch and discuss the verification test videos in a MIV 2 BoG, especially the best reference and source views.
- It is encouraged to bring MIV demos and show them to non-MIV MPEG experts

3.1.2 April 23

- The target is to have an WG04 output document of this meeting with the final verification test report
 - Review of the draft of the verification report
- We want to remove Balloons because the quality could not be enhanced
- Mathias has received the document and asked for some time to review

3.2 m62490 [MIV] Improvement of S/F sequences rendering quality on patch loss

- The proponent has performed the experiment using TMIV 15 and the new CTC.

- [@bartkroon](#) volunteers to perform a crosscheck. [@yuxiaobai](#) will provide the relevant information.
- The FTP service is down for most of us and it is not possible to upload content. This may be security related.
- The proponent may use other means to share the data with MIV experts.

This document was replaced by m63213

3.3 m62529 Draft text for ISO/IEC FDIS 23090-23

Authors state that with proposed changes it is safe to go to FDIS.

3.4 m62575 Summary of Voting on ISO/IEC DIS 23090-23

In this MPEG meeting, we need to:

- Output DoC
- Output FDIS text

3.5 m62701 Adaptive patch-wise depth range linear scaling

- Applying the per-patch depth range to all patches performed better, then applying it only on larger ones
- Crosscheck was successful (see [#683](#))
- [@bartkroon](#) (MIV editor + SW coordinator) want to study the source code, especially TMIV decoder-related parts to make sure that this change matches with the MIV 1 specification
- There is in general support for per-patch depth ranges
- We want to watch pose trace videos when available, but we believe that we already have sufficient evidence.
- Q: Does the patch offset and patch scaling impact each other? A: Independent, and compatible.

[@sg.lim](#) [@adziembowski](#) and [@bartkroon](#) will discuss offline how to combine in a MIV 1 compatible way.

3.6 m62894 On default values for some MIV syntax elements

Adopted. The contribution has a choice. For that there is a preference to block the `aaps_vpcc_extension_enabled_flag`.

Discussion: for a new MIV profile it would (of course) be possible to use the AAPS extension because that new profile may not block that flag.

3.7 m63015 New depth maps for Breakfast sequence

- It was difficult to judge the pose traces because of the network connection.

- We can watch pose trace videos in the AG 05 room sometime mid week.
- [@bartkroon](#) will ask [@yulu](#) if a solution can be found for the MPEG content server.
- Fall back is to share a writable temporary FTP for MIV experts only.

We will revisit this proposal after MIV experts had the opportunity to study the material in more detail.

3.8 m63047 [MIV] Demonstration of the bullet time concept

- Demo without encoding, just rendering of bullet-time-like video to promote the concept.
- Demo allows to remove objects which are separated from the background.

3.9 m63048 [MIV] Summary Report on EE7(Extended entity-based coding)

There is a problem with the sub-picture merging tool and the VTM software coordinator does not know if that developer is still available

It is proposed to continue EE7 in a different direction, to improve entity-based coding. There is a related document that will be presented.

3.10 m63051 [EE-7-related] Spatio-temporal merge (STM) of background object for MIV v2

- Foreground and background is separated
- Temporal merge to inpaint background from other frames in the GOP
- Spatial merge to synthesize a central background view; 1 intra frame BG for 32 frames; in one intra period
- Spatio-temporal merge does both
- Tests were run on Museum for which entity ID maps are available
- CTC anchor outperforms separated FG/BG anchor that was created in this experiment

Discussion:

- It is surprising that spatial merge works better than temporal merge; spatial merge objectively also outperforms spatio-temporal merge

Two possible reasons for the Spatio-temporal mode (STM) does not perform better than the Spatial mode (SM):

1. Background generated from STM contains the future data as well, which is better because we have less empty space. But the current PSNR computation does not reflect this fact.
2. Compression performance is not that different between STM and SM, since the inter-frame video coding works quite well on background temporal frames so tha overall bit rate difference is very marginal.

Objective performance does not benefit but subjective evaluation could because of the inpainting of the disoccluded regions.

- It is important to watch pose trace videos.
- The splitting in foreground/background causes a BD-rate reduction.
 - The bit rate increase can be explained by the additional information: is a pixel foreground or background?
 - The quality decrease requires investigation according to the proponent. It is encouraged to perform a rendering test with the best reference vs. rendering the all entity layers separately. One way to do this, is by splitting views in foreground views and background views, and render those separately using the best reference.

The proposal needs revision because of the quality degradation.

The AHG recommends to continue EE7

3.11 m63058 [MIV] [V-PCC] Real-time decoding and rendering demo on a smart phone

- Demo showing virtual navigation done using phone's gyroscope or finger movement.
- Two new sequences from Philips (Soccer - 8 cameras, Dance - 6), shot with help of professional company, processed and coded by internal Philips tools.
- Point-cloud and MPI sequences were also shown.
- Demo focuses on maximizing quality, tools that can deteriorate the quality were turned off (e.g. splitting of patches was not allowed).
- Coding compatible with MIV (without newest conformance fixes), 1 atlas, 4k x 3k resolution, 21 MBit/s, h265 with high preset. CTC seems to be a bit too liberal with its current restrictions. Pixel rate is much bigger problem than bit rate at the moment.
- Quality shown to be good, only minor rendering issues were spotted.

3.12 m63109 Renderer-side edge processing for subjective quality improvement

- The viewport is filtered to filter down edges at depth jumps
- The purpose is to make the viewport look more natural
- No objective evaluation was performed
- The method was tried in the context of VT preparations
- In that context a subjective improvement was observed by multiple organizations

Q: Why only on the viewport? Why not on all patches? A: Difficult with the current renderer, but could improve overall rendering. Difficult to expect.

The AHG recommends to integrate this method into TMIV 16 and enable in the CTC. The threshold will be set in the JSON file (TMIV decoder configuration).

3.13 m63112 Proposal of IVDE 8.0

- Skipping of null optimizations (empty volume layers) to reduce runtime complexity
- Only compute sub-pixels in neighboring layers for subsequent passes

- Two round-related bugs were fixed; they probably had only a minor performance impact
- It is proposed to go back to 1 thread
- Pose traces are available

3.14 m63113 Patch geometry offset modification

- Syntax to enable geometry offset
- AN experiment was performed
- Y-PSNR improves more than IV-PSNR because IV-PSNR allows for a small block displacement
- There is an objective improvement on average
- The proponent does not expect that the pose traces will have a subjective improvement
- The proposed syntax is modelled after the pdu_texture_offset HLS.

Discussion:

- It is argued that existing PDU syntax is enough
- The hypothetical reference renderer describes how to apply this
- This was implemented in the TMIV decoder but incorrectly, see bug <https://mpeg.expert/software/MPEG/MIV/RS/TM1/-/issues/745>
- The proponent indicates that there appears to be a problem with the semantics

3.15 m63212 Proposal of a New Multi-ToF Natural Content: CoffeeTime

- Recording with eight Kinect cameras
- Three computers were used to capture: server has 4 cameras, one computer had 3 cameras and one computer had 1 more cameras
- License Attribute-NonCommercial-ShareALike 4.0 International
 - This license is suitable for the promulgation of the standard, as long as no money is made because of the NonCommercial clause
- Q: Rendering? A: Not yet
- The content is available (see above) if people want to try
- [@dmieloch](#) wants to try to use this sequence with IVDE, to use IVDE to fill in the missing pixels
- The proponent ([@lafruit](#)) remarks that this content is meant for MIV 2 and it would be good if the TMIV decoder/renderer can handle these kind of depth maps even with holes.
- [@dmieloch](#) / [@adziembowski](#), with help of the proponents, will convert to normalized disparity, and then this can also be used for direct coding with MIV for comparison.
- [@Gwangsoon](#) reminds us that there is also a coloured depth use case/requirement for MIV 2.
- No technology was provided so far for coloured depth

It is recommended to use this content in MIV 2 related studies.

3.16 m63213 [MIV] EE8 report: Encoder-side Effective Information (ESEI) Based optimization of multi-view atlas generation

- Compared to the previous EE8-related contribution, splitting of clusters based on effective information was added
- The cluster is split to maximize the difference in effective information

There is support for this work. We want to see some pose traces and a crosscheck is needed. [@Gwangsoon](#) offered to provide a crosscheck.

3.17 m63219 [MIV] Prototype MIV player for demonstration

- Player based on TMIV 13.1, with CUDA for video decoding and rendering. OpenGL was only used for rendering the quad to screen.
- Two sequences were tested: Cyberpunk and Breakfast
- Three encoder modes were tested:
 - MIV Main with only basic (full) views,
 - MIV Main with basic and additional views (with patches)
 - MIV Main with basic and additional views with reduced views and one atlas
- Rendering speed between 13 - 50 fps depending on the condition

It would be good to view all demos again later in the week and give more people the opportunity to see them.

3.18 m63396 [MIV] Crosscheck of m62701

The crosscheck was successful.

3.19 m63397 Chroma dynamic range modification

- Loss for Y-PSNR because more bits are used for chroma
- Gain for IV-PSNR because chroma was coded more accurately
- It was not yet possible to provide pose traces due to problems with the MPEG content server

Discussion:

- With mobile implementations often a texture in RGB format is received, and it is difficult to apply a filtering of YUV values in that case.
- There could have been a flag in the VPS to indicate that texture offsets are enabled
- There is a flag in the CASPS
- Q: Can this be done in a different way? A: With setting different QP's for luma and chroma; but this is limited by a difference of 12.
- Q: Why does chroma needs so much better with MIV? A: Normally chroma does not matter much, but with the patch edges the chroma difference is visible.

- In JVET there is an activity relating to patch video coding; Sebastian Schwarz is involved; that activity may be interested in this problem. The same problem may also occur with V-PCC.
- The proponent argues that changing the video encoder will not solve the issue because it is not possible to take action on a patch level.
- The proponent argues that the only visible patch edge artifacts are now due to luma, but not chroma
- The same method may not work as good with luma because luma more typically has a larger range and thus cannot be expanded that far

Revisit in a BoG call.

4 AHG recommendations

Adopted into ISO/IEC 23090-12 MIV edition-2:

- m62894 On default values for some MIV syntax elements

Adopted into ISO/IEC 23090-23:

- m62529 Draft text for ISO/IEC FDIS 23090-23

Integrate into TMIV 16:

- m63109 Renderer-side edge processing for subjective quality improvement

Appreciate UPM and ULB (full names) for providing the CoffeeTime sequence:

- m63212 Proposal of a New Multi-ToF Natural Content: CoffeeTime

View all demos again:

- m63219 [MIV] Prototype MIV player for demonstration
- m63047 [MIV] Demonstration of the bullet time concept
- m63058 [MIV] [V-PCC] Real-time decoding and rendering demo on a smart phone

View pose trace videos of:

- m62701 Adaptive patch-wise depth range linear scaling
- m63015 New depth maps for Breakfast sequence
- m63112 Proposal of IVDE 8.0

Review software of:

- m62701 Adaptive patch-wise depth range linear scaling

Revise documents:

- m63051 [EE-7-related] Spatio-temporal merge (STM) of background object for MIV v2
- m63113 Patch geometry offset modification
- m63213 [MIV] EE8 report: Encoder-side Effective Information (ESEI) Based optimization of multi-view atlas generation