 ISO/IEC JTC 1/SC 29/WG 04 N0054

**ISO/IEC JTC 1/SC 29/WG 04  
MPEG Video Coding   
Convenorship: CN**

**Document type:** Output Document

**Title:** Call for MPEG-I Visual Test Materials

**Status:** Approved

**Date of document:** 2021-01-22

**Source:** ISO/IEC JTC 1/SC 29/WG 04

# Expected action: None

# Action due date: None

**No. of pages:** 5

**Email of Convenor:** yul@zju.edu.cn

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg4>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 04 MPEG VIDEO CODING**

**ISO/IEC JTC 1/SC 29/WG 04 N** **0054**

**January 2021, Online**

|  |  |
| --- | --- |
| **Title** | **Call for MPEG-I Visual Test Materials** |
| **Source** | **WG 04, MPEG Video Coding** |
| **Status** | **Approved** |
| **Serial Number** | **20006** |

**Introduction**

During exploration experiments, MPEG has realized that there is a lack of test material for video experiments into MPEG-I Immersive Video [N0065]. This is especially true for the 6DoF and Dense Light Field processing and coding experiments.

Understanding this need, we hereby solicit new test material for 6DoF and Dense Light Field, **in particular using non-parallel camera setups**. Since all technologies in MPEG-I Visual are based on multiple camera views and their depth maps, we recommend that not only the camera views, but also depth maps (measured or estimated at high quality) are provided by the proponents. Furthermore, in order to prepare for more advanced features of next generation Immersive Video, capture/creation of **non-Lambertian content** with specular reflections, transparent objects are also encouraged.

Content may be provided as computer-generated/synthetic **3D models of dynamic scenes**, as this material can be used for rendering various viewpoints with computer graphics techniques, creating video footage required in all experiments and comparative studies (e.g. PSNR quality evaluation with reference views).

MPEG is also calling for **natural content**, both **indoor and outdoor**, directly captured with camera rigs. Content with **objects close to the camera** are also requested, since this will challenge the proposed technologies for parallax rendering, e.g. heavy motion parallax for nearby foreground objects.

Content should be provided in any **image-based representation format**, e.g. **lenslet** format, or **multiview+depth**. If content is generated in computer-generated/synthetic format, it should be rendered to create various image-based projections (and their depth maps) to be eligible as test material.

Please consult [N0053] for an overview of available test material.

**MPEG Immersive video content**

MPEG Immersive video [N0065] enables coding a volumetric scene captured from a multitude of cameras in a regular or irregular arrangement. Inter-camera baseline distances can vary and may be narrow or wide. The cameras provide their captured videos in a known projection format; for e.g. perspective, ERP. An MIV decoder and renderer, given an MIV coded bitstream, can render complete views for any viewpoint that is within a specified viewing volume.

The purpose is to prepare content for MIV core experiments and 6DoF exploration experiments. Content will be used to define pose traces and common test conditions for evaluating competing coding technologies, as well as pre-processing (e.g. depth estimation) and post-processing (e.g. rendering by view synthesis) algorithms.

Typical content that would be required with high priority are:

* Multiview+Depth, with increased viewing volume dimension, more cameras, larger distance between cameras than current content used for MIV.
* High quality depth maps (provide confidence map, if available).
* Inward looking and outward looking camera setups, with arbitrary configuration possible (includes any non-trivial configuration).
* Natural content and outdoor scenes.

Content with following properties is highly encouraged for preparing the second generation MIV and 6DoF experiments, cf. [N0055]:

* Any projection type corresponding to a physical camera rig (e.g. perspective, fisheye).
* Fine geometry.
* Dynamic scenes with larger viewing volumes and more cameras.
* Complex light interactions within the scene.
* Scenes with basic transparencies introduced as RGBA images completed by fully opaque views to be used as reference source views.
* Deep image (such as OpenEXR) or RGBA format, with multiple attributes per pixel beyond geometry and texture, such as reflection, refraction, transparency, object, and material ID.
* Moving cameras with step-in/step-out motion and/or change in view orientation.
* Scenes with particle mediums such as fog, cloud, water.
* Non-static content, with biological entities (for e.g. people, cats, dogs, grass, hair/fur).

Camera intrinsics and extrinsics should be provided in json format, cf. [N0050].

It would be convenient, if the textures, depth and attributes are provided in a YUV format. If other formats are provided, it would be highly desirable that a converter to convert from the proposed format to a YUV format also be provided.

Additionally, we are also soliciting video content in the form of Multi-Plane images (MPI) ([MPIGoogle]). This representation consists of an ordered set of layers, each layer representing texture and transparency with a constant depth. The depth is itself ideally represented as a normalized disparity value. It is also possible to provide multiple MPI sequences for the same volumetric scene, for example MPIs generated from different viewing position and orientation. The MPI sequences can be provided as either one of two alternates:

* as YUV texture plus transparency format, or
* as files in OpenEXR 2.0 format.

The YUV texture plus transparency format consists of one yuv420p10le YUV sequence file for texture and another yuv400p8 YUV sequence file for transparency. Each frame of the MPI sequence (both texture and transparency) is a concatenated sequence of MPI layers ordered by decreasing depth (furthest first to closest last). An MPI sequence, for both texture and transparency, is a concatenation of MPI frames in presentation order. A JSON file should include all necessary information, and a document describing the process of the depth value recovery of each layer should also be provided.

As an example, the JSON file can include the total number of layers *n*, the parameters of the camera used for the MPI including the minimum disparity value, 1/*zmin*, and maximum disparity value, 1/*zmax*, the number of bits, *b*, used for coding the depth values.

Disparity, *d*, is then computed by the equation

*d* = *l* + (*h* – *l*) \* *b* / (2\*\**n* – 1),

where *d* denotes disparity of a layer, *h* denotes the maximum disparity, *l* denotes the minimum disparity, *b* denotes the number of bits for coding the disparity, and *n* denotes the total number of layers.

A second option is to use OpenEXR 2.0 [OPENEXR20] which allows specifying a variable list of samples per pixels, to generate deep images. multiple values at different depths could be stored for each pixel in the sequence. Additionally, each file can contain a number of separate, but related, data parts. This enables storing of multiple MPIs in a single file, if needed. The compression of the OpenEXR files should be chosen judiciously to avoid artefacts.

**Dense Light field content**

Dense Light Fields can either be captured by dense multi-camera arrays or lenslet cameras (e.g. Lytro or Raytrix). Currently, two formats: multiview and lenslet, are commonly used. These data formats can be converted from one to the other. However, such a conversion may or may not be invertible, depending on whether a plenoptic 1.0 or a plenoptic 2.0 camera model is used.

The Dense Light Field activity in WG-04 of MPEG aims to compare the coding performance of different potential representations of dense light field content (lenslet, multiview, etc), and explore new application scenarios in immersive video. Currently, there are more test sequence in dense multiview format than those using plenoptic camera. We solicit, new lenslet video content in plenoptic 1.0 and 2.0 format, with a priority for the latter.

Furthermore, MPEG has initiated exploration on new applications using dense light field videos obtained from plenoptic camera arrays for immersive video. Therefore, we solicit new test content as input to the MPEG-I-Visual AHG, following the description below.

1. Multiview lenslet video captured by an array of plenoptic cameras,
2. Dense multiview video corresponding to each lenslet video in (1), and/or the conversion tool to synthesise such dense multiview video,
3. Multiview depth map video corresponding to the dense multiview video in (2), and/or the tool to estimate such depth maps, and
4. Camera parameters of the plenoptic cameras in (1), and corresponding virtual cameras of the dense multiview video in (2).

**3D synthetic models**

In the case of photorealistic synthetic content, it is recommended that complementary to the rendered image data, also the 3D model itself plus any scripting to generate the required dataset is supplied. This enables other MPEG experts to render variants of the same scene. It is for instance possible to simulate a virtual camera array by placing multiple cameras in the scene. Preferably, scenes should include a timeline that allows for rendering of short movies with some dynamics. Due to the open source and easy accessibility of the Blender software, a blend file would be the most preferred format for this type of content.

The format of the 3D model should be in a suitable interchange format [N17252], or open source (e.g. Blender[[1]](#footnote-1)). Typical representations of 3D models are:

* Texture, mesh and lighting information,
* Procedural scenes (e.g. algorithms).

Content providers are nevertheless asked to render their 3D scene in order to create multiview image-based representations. Also, the depth maps should be created in the rendering process.

**Experimentation test set**

In addition to photorealistic synthetic content that is solicited, simple Cornell-Box type scenes as illustrated in Figure 1 are also solicited. These types of focused scenes are useful to test individual aspects of a volumetric video and the experimentation and evaluation of algorithms would become easier.

|  |  |
| --- | --- |
| A picture containing wall, tree, indoor, green  Description automatically generated | A picture containing indoor, wall, green, ceiling  Description automatically generated |

**Figure 1: (left) a Cornell-Box with a smoke simulation; (right) a Cornell-Box with a pure reflective cuboid and a glossy sphere.**

**Copyright notice**

Content owners should provide a copyright notice along with the dataset to inform MPEG about copyright and usage restrictions.

**Contact:**

* [Vinod Kumar Malamal Vadakital](mailto:vinod.malamalvadakital@nokia.com)
* [Renaud Dore](mailto:Renaud.Dore@InterDigital.com)

**References**

[N0065] Text of ISO/IEC DIS 23090-5 Visual Volumetric Video-based Coding and Video-based Point Cloud Compression 2nd Edition, ISO/IEC JTC1/SC29/WG04 MPEG2020/[N0065](http://phenix.int-evry.fr/mpeg/doc_end_user/current_document.php?id=74671&id_meeting=), Online. Jan. 2021.

[N0050] Test Model 8 for MPEG Immersive Video, ISO/IEC JTC1/SC29/WG04 MPEG2020/N0050, Online, Jan 2021.

[N0053] Overview of MPEG-I Visual Test Materials, ISO/IEC JTC1/SC29/WG04 MPEG2020/N0053, Online, Jan 2021.

[N0055] Exploration Experiments on Future MPEG Immersive Video, ISO/IEC JTC1/SC29/WG04 MPEG2020/ N0055, Online, Jan 2021.

[N17252] Interchangeable Formats in VFX Industry (v1), ISO/IEC JTC1/SC29/WG11 MPEG2018/N17252, Macau, Oct 2018.

[N18793] Manual of Reference Lenslet Content Convertor (RLC 1.0), ISO/IEC JTC1/SC29/WG11 MPEG2019/N18793, Geneva, Oct 2019.

[MPIGoogle] Immersive Light Field Video with a layered mesh representation, Broxton et all, SIGGRAPH 2020 Technical Paper, <https://augmentedperception.github.io/deepviewvideo/>

[OPENEXR20] <https://www.openexr.com/>

1. https://www.blender.org/ [↑](#footnote-ref-1)