**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2018/N17608**

**January 2018, Gwangju, Korea**

|  |  |
| --- | --- |
| **Source** | **Video** |
| **Status** | **Approved** |
| **Title** | **Call for MPEG-I Visual Test Materials on 6DoF** |
| **Editors** | Bart Kroon, Gun Bang, Mehrdad Panahpour Tehrani, Gauthier Lafruit |

# Introduction

During exploration experiments, MPEG has realized that there is a lack of test material for video experiments into MPEG-I Immersive Video [[N17685](http://wg11.sc29.org/doc_end_user/current_document.php?id=62111&id_meeting=174)], especially for 3DoF+ and Omnidirectional 6DoF, but also for Windowed 6DoF and Dense Light Field.

We hereby solicit new test material for all categories. 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) are provided by the proponents.

Content may be provided as computer-generated/synthetic **3D models of static and 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 warmly recommended, 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 [[N17606](http://wg11.sc29.org/doc_end_user/current_document.php?id=62032&id_meeting=174)] for an overview of available test material.

# Omnidirectional 6DoF

Recent contributions have brought new interesting (synthetic) content reaching photorealism with perfect depth maps, e.g. [[m41824](http://wg11.sc29.org/doc_end_user/current_document.php?id=60538&id_meeting=172)], unfortunately only for static content. New omnidirectional - and preferably natural - test sequences with good depth maps are requested to conduct the DIBR virtual view synthesis as described in EE\_Synthesis [[N17606](http://wg11.sc29.org/doc_end_user/current_document.php?id=62032&id_meeting=174)].

According to the the MPEG-I PDTR [[N17685](http://wg11.sc29.org/doc_end_user/current_document.php?id=62111&id_meeting=174)], Omnidirectional 6DoF corresponds to a viewer being able to take multiple steps in the virtual scene. Therefore, we request content of any (virtual) multi-camera configuration with an omnidirectional nature (at least 180°) and nearby objects. The content should be provided as the YUV 420 or YUV 444 uncompressed video output of each camera in the rig. Examples of potentially suitable rigs can be Nokia Ozo, Facebook Surround 360, Vuze or any other custom-built rig. These camera systems typically have ultra-wide-angle or fisheye lenses. Intrinsic and extrinsic camera parameters have to be included, and it is highly recommended, but not strictly required, for a participant to supply depth maps or scripts to generate them. If the depth maps are provided, it should be in the YUV 400 video format.

In case of ultra-wide-angle (e.g. 140°) and fish eye lenses, we ***recommend*** receiving both the original and rectified video frames. For camera frames with a smaller FOV, rectified video frames are sufficient.

# Windowed 6DoF

Windowed 6DoF requests content obtained in a convergent way, where the set of cameras are outside the scene that is intended to be rendered. The goal is to enable some virtual navigation in the scene.

MPEG appreciates both natural captured content and computer generated one. The content shall be provided together with corresponding depth maps (captured or estimated). The camera parameters (intrinsic and extrinsic) shall be provided only for texture if depth and texture have the same optical center, otherwise camera parameters shall be provided both for texture and depth. The content shall be provided in YUV 420 or YUV 444 formats for texture and in YUV 400 for depth maps. Depth information may be characterized using the depth formats described in [N16730]. The camera array should have a 2D arrangement following as much as possible a planar, cylindrical or spherical surface, though any camera arrangement is eligible. 1D arrays with a large number of views are also highly appreciated. MPEG encourages proponents to provide video material with framerates larger than 25 fps.

# Dense representation of Light fields

Dense Light Fields can be either captured by dense multi-camera arrays or a lenslet camera, e.g. Lytro or Raytrix. Currently, two formats (multiview and lenslet) are considered. These data formats can be converted from one to another. Such a conversion can be lossy or lossless.

This activity aims to compare the coding performance of different potential representations of dense Light Field data (lenslet, multiview, etc). Currently, several test sequences with dense multiview video [ULB Unicorn, Champagne\_tower, Pantomime, Big Buck Bunny, Ostendo], and only a couple with lenslet [Nagoya m41787, m41995: Tunnel\_Train\_1 and 2, INRIA m42468] are provided.

We hence encourage participants to provide MPEG test material for this activity with following contents, by MPEG123: (1) lenslet video, dense multiview video (2) multiview/lenslet converted from lenslet/multiview, (3) camera parameters, and (4) the conversion tool for conversion from lenslet to multiview video and vice versa.

# 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.

The format of the 3D model should be in a suitable interchange format [[N17252](http://wg11.sc29.org/doc_end_user/current_document.php?id=60789&id_meeting=172)], 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 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.

**Contact:**

* Bart Kroon, [bart.kroon@philips.com](mailto:bart.kroon@philips.com)

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