**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N16766**

**April 2017, Hobart, Australia**

|  |  |
| --- | --- |
| **Source** | **Requirements** |
| **Status** | **Approved** |
| **Title** | **Call for immersive visual test material** |
| **Editors** | Krzysztof Wegner, Gauthier Lafruit |

# Introduction

# MPEG started to work on the new work item MPEG-I targeting future immersive applications. In particular, this new standard will enable various forms of audio-visual immersion, including panoramic video with 2D and 3D audio with various degrees of true 3D audio-visual perception.

MPEG is calling for video test material to assess algorithm performance for different setups where information is combined from different cameras to generate virtual views that the user observes when virtually walking in a scene. Different levels of experience are achieved by the user who may freely move his head around three rotational axes, aka Three Degrees-of-Freedom (3DoF), and along three translational directions (including stepping forward and backward into/from the scene), aka Six Degrees-of-Freedom (6-DoF). In order to optimize the standard for the intended applications, MPEG requests interested parties to submit test material to MPEG.

# Test material characteristics

Test material for the development of the next-generation, immersive video coding standards should comply to the attributes described below.

## General considerations

### Still image and video sequences

Both still image and video sequences are solicited.

High SLR resolutions are recommended, i.e. at least 1920x1080 HD resolution; higher resolutions are preferred, e.g. 20 Mpixels or more. Video material should have frame rates of at least 25 fps.

Test material from both indoor and outdoor scenes can be submitted, with sufficient complexity to test the limits of the algorithms under study. For example, content with fur, hair, transparent objects, specular regions, etc. is recommended.

### Natural vs. Computer-generated content

Natural content is highly preferred over computer-generated content. Depth material for all content is solicited, if available, using depth sensing devices and/or depth estimation techniques.

### Color components, depth, and metadata

In general, it is required that texture, depth, and metadata are provided separately, in particular for the camera parameters. All texture content should be provided in the three-channel format of RGB, YUV444 or YUV420. Depth information, if available, should be provided. Depth information may be characterized using the depth formats described in [N16730].

### Types of cameras and camera array arrangements

The types of cameras that are envisioned to be relevant for this call include:

1. Cameras that capture a highly dense array of images along a predefined track (e.g. 2D linear with parallel cameras, 2D linear with convergent cameras, 2D cylindrical surface, 2D spherical surface) thereby producing the equivalent of a fixed camera array with either convergent or divergent views.
2. Regularly spaced, fixed camera arrays of at least 10 cameras arranged in 2 or more rows. Arrays of higher vertical density are preferred (to support step-in/out functionality studies).
3. Plenoptic (e.g. lenslet) cameras, or other cameras where a microlens array is inserted between the main lens and the image sensor.
4. Divergent camera arrangements of at least 6 cameras with significant overlapping fields of view, and the ability to provide the content in an unstitched, raw form.

The type of cameras used should be reported with data sheets. In particular, for video it’s important to report whether a global or rolling shutter camera type is used. Information about extrinsic and intrinsic properties of the devices is required to support the MPEG-I activities. Their format should be provided as described in Annex 1. Even if the views are captured over a regular grid, additional registration (alignment to a perfectly regular grid in post-processing) is preferred. Color calibration over all views, reaching the same color tones over all views, is desired.

### Synchronization

Accurate temporal synchronization of multiple cameras is preferred. This means that all cameras should take images at the same time with an external global clock signal. Alternatively, manual or post-production time alignment shall be applied. Global shutter cameras are recommended in video acquisitions.

## Omnidirectional video with depth data

The content should be captured with an arrangement of cameras that records divergent views, preferably in an arrangement that supports the capture of a full 360-degree field of view. This, for example, may be an array of cameras that are arranged on the surface of a sphere. Both the texture and depth data must be provided at the same resolution with an input greater than or equal to 4K, and the same projection; preferably in the equirectangular projection.

## Video material recorded by divergent camera arrangement with significant overlap

The content should be captured with an arrangement of cameras that record divergent views, preferably in an arrangement that supports the capture of a full 360-degree field of view. This, for example, may be an array of cameras that are arranged on the surface of a sphere. In this case, the captured content must be raw and unstitched data. For this scenario, both the intrinsic and extrinsic camera parameters must be provided.

## Video material recorded by convergent camera arrangement with significant overlap

The content should be captured with an arrangement of cameras that record convergent views, preferably in an arrangement that supports the capture of a volume of visual data. This, for example, may be an array of cameras that are arranged in the shape of a dome (e.g. hemisphere). In this case, the raw texture data must be provided. Additionally, the extracted point cloud representation is appreciated. Both the intrinsic and extrinsic camera parameters must also be provided.

Static material recorded by ultra-dense camera arrangement that supports step-in/step-out exploration activities are solicited.

## 2D camera array arrangement

The cameras should be in a 2D arrangement, following a planar, cylindrical or spherical surface. Dense video sequences are particularly sought with a baseline distance between cameras not more than 20cm, and a total baseline (the distance from one end of the array to the other end) as wide as possible. In this scenario, depth maps for each of the cameras must be provided, preferably at the same resolution as the texture data. Both the intrinsic and extrinsic camera parameters must also be provided for all the cameras. If the depth data are provided at spatial positions that are different from the camera views, then the intrinsic and extrinsic parameters of the depth sensor must also be provided. If rectification is necessary to provide views in linear and parallel arrangement, it is recommended to provide both original and rectified data. Rectification should be done by using high quality algorithms.

## Plenoptic cameras

Plenoptic cameras are defined as acquisition systems including an array of micro-lenses between the main lens and the sensor. The density of micro-lenses is supposed to be large enough to ensure a good angular sampling of the light field. Resolution of the plenoptic image should be no less than 15 mega-rays.

High resolution depth maps (at the same spatial resolution as the RGB data) are desired.

For viewing the content, a suggested player that is available free-of-charge is asked to be identified for each submission. Please specify the system requirements that are required for player executables.

## Systems of simultaneous multiple acquisitions

For this scenario, more than one of the acquisition systems defined above is considered. These systems shall simultaneously acquire the same scene. Requirements for each of the acquisition systems used should follow the specifications defined above (e.g. camera parameters, resolutions).

Extrinsic parameters of the entire system are required for the MPEG-I activities.

# Contacts and other logistics

For questions or to respond to this call, please contact:

Masayuki Tanimoto [tanimototentative3@yahoo.co.jp](mailto:tanimototentative3@yahoo.co.jp)

Krzystof Wegner: [kwegner@multimedia.edu.pl](mailto:kwegner@multimedia.edu.pl)

Gauthier Lafruit: [gauthier.lafruit@ulb.ac.be](mailto:gauthier.lafruit@ulb.ac.be)

Submissions in response to this call are kindly asked to be made by the 119th MPEG meeting which is planned for 17-21 July 2017 in Torino.

# References

[N9595] Call for Contributions on 3D Video Test Material, ISO/IEC JTC1/SC29 WG11 N9595, Antalya, Turkey, January 2008

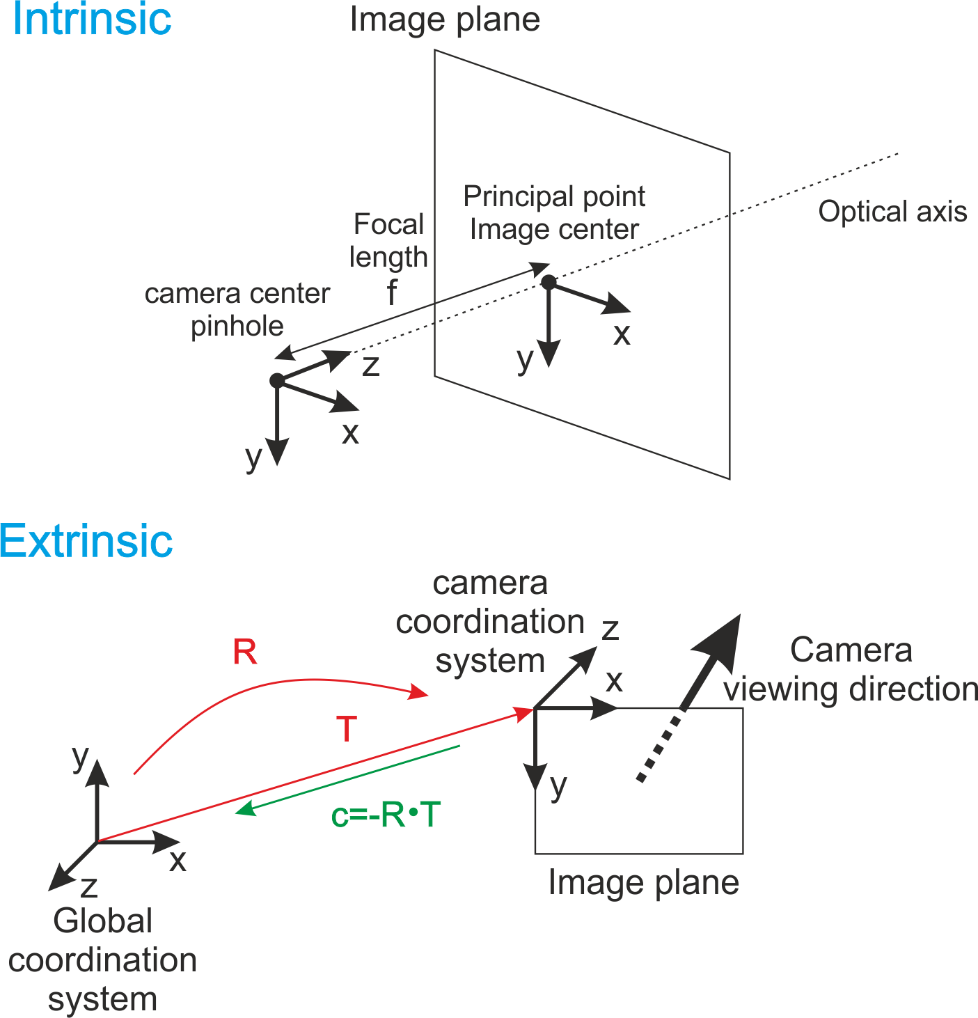
[N16730] Depth map formats used within MPEG 3D technologies, ISO/IEC JTC1/SC29 WG11 N16730, Geneva, Switzerland, January 2017

# Annex 1: Camera parameter formats

Camera parameters principles

The camera parameters are subdivided into two categories:

* the camera intrinsics, i.e. the parameters of each camera individually: camera center c position in world coordinate system, principal point *p* and focal length *f* in pixel units, as shown in the top part of Figure 1.
* the camera extrinsics, i.e. the camera rotation R and translation T, expressed in world coordinates, as shown in the bottom part of Figure 1.



**Figure 1: Intrinsics and extrinsics of cameras**

Literature often uses the [R|c] matrix at the bottom of Figure 1 for the extrinsics. However, MPEG’s depth estimation and view synthesis software uses a slightly different format, replacing c by the camera position T = - R-1 \* c, according to [N9595], which annex A is copied in the next section for completeness.

Specification of Camera Parameters

Camera parameters shall be specified as rotation matrix **R**, translation vector **t** and intrinsic matrix **A** for each camera i. Values shall be given in floating point precision as accurate as possible.

The extrinsic camera parameters **R** and **t** shall be specified according to a right-handed coordinate system. The upper left corner of an image shall be the origin for corresponding image/camera coordinates, i.e., the (0,0) coordinate, with all other corners of the image having non-negative coordinates. With these specifications, a 3-dimensional world point, **wp**=[x y z]T is mapped into a 2-dimensional camera point, **cp** = s \* [u v 1] T, for the i-th camera according to:

s \* **cp**(i) = **A**(i) \* **R**-1(i) \* [**wp** – **t**(i)]

where **A**(i) denotes the intrinsic camera parameters, **R**-1(i) denotes the inverse of the rotation matrix **R**(i) and s is an arbitrary scaling chosen to make the third coordinate of **cp** equal to one.

The rotation matrix **R**(i) for i-th camera is represented as follows.

|  |  |  |
| --- | --- | --- |
| **r\_11**[i] | **r\_12**[i] | **r\_13**[i] |
| **r\_21**[i] | **r\_22**[i] | **r\_23**[i] |
| **r\_31**[i] | **r\_32**[i] | **r\_33**[i] |

The translation vector **t**(i) for i-th camera is represented as follows:

|  |
| --- |
| **t\_1**[i] |
| **t\_2**[i] |
| **t\_3**[i] |

The rotation matrix **R** and the translation vector **t** define the position and orientation of the corresponding camera with respect to the world coordinate system. The components of the rotation matrix **R** are function of the rotations about the three coordinate axes.

**focal\_length\_x**[i] specifies the focal length of the i-th camera in the horizontal direction units of pixels.

**focal\_length\_y**[i] specifies the focal length of the i-th camera in the vertical direction in units of pixels.

**principal\_point\_x**[i] specifies the principal point of the i-th camera in the horizontal direction units of pixels.

**principal\_point\_y**[i] specifies the principal point of the i-th camera in the vertical direction in units of pixels.

**radial\_distortion**[i] specifies the radial distortion coefficient of the i-th camera.

The intrinsic matrix **A**(i) for i-th camera is represented as follows:

|  |  |  |
| --- | --- | --- |
| **focal\_length\_x**[i] | **radial\_distortion**[i] | **principal\_point\_x**[i] |
| 0 | **focal\_length\_y**[i] | **principal\_point\_y**[i] |
| 0 | 0 | 1 |