**ISO/IEC JTC 1/SC 29/WG 11  
Coding of moving pictures and audio  
Convenorship: UNI (Italy)**

**Document type: Approved WG 11 document**

**Title: Description of Core Experiment on Immersive Media Metadata**

**Status: Approved**

**Date of document: 2020-02-28**

**Source: Systems**

**Expected action:**

**No. of pages: 7**

**Email of convenor: leonardo@chiariglione.org**

**Committee URL: http://mpeg.chiariglione.org**

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/WG 11 N19033**

**Brussels, BE – January 2020**

**Source: Systems**

**Title: Description of Core Experiment on Immersive Media Metadata**

# Introduction

Several use cases requiring metadata to indicate the relationship between 2D pixels in a coding domain and 3D volumetric representation including support of efficient processing, annotation and so on have been identified. Systems has been working on a CE defining some metadata for identifying and delivering some subset of data to improve efficiency of processing until the 129th MPEG meeting. At the 129th MPEG meeting, it has been decided to modify the scope of such CE to consider other use cases for the metadata and find harmonized solution.

Strong consensus has been built over the last several meetings that for many application entire PC object data does not have to be delivered, decoded and rendered as only some portion of a PC object would be presented to the user at once. Many solutions to efficiently support such scenario have been proposed as well. To facilitate comparison of proposed technologies, selection of appropriate technologies and design of a harmonized solution if needed, a core experiment has been established at the 127th MPEG meeting. Throughout the CE process, a basic framework enabling partial access of a PC object has been agreed and implemented in the revised CD (N18832). The collection of proposed technologies still under evaluation are currently documented in Technologies under consideration on carriage of PC data (N18833). This document describes details of the CE to be performed until the 129th MPEG meeting to evaluate additional technologies to enhance this feature, e.g. spherical subdivision of a PC object.

# Logistics

## Coordinator

Youngkwon Lim (yklwhite@gmail.com)

## Participants

* Apple
* Huawei
* Intel
* InterDigital Communications, Inc.
* Kyunghee University
* LG Electronics Inc.
* MediaTek
* Nokia
* Samsung
* Sharp
* Sony
* Tencent
* ZTE

## Schedule

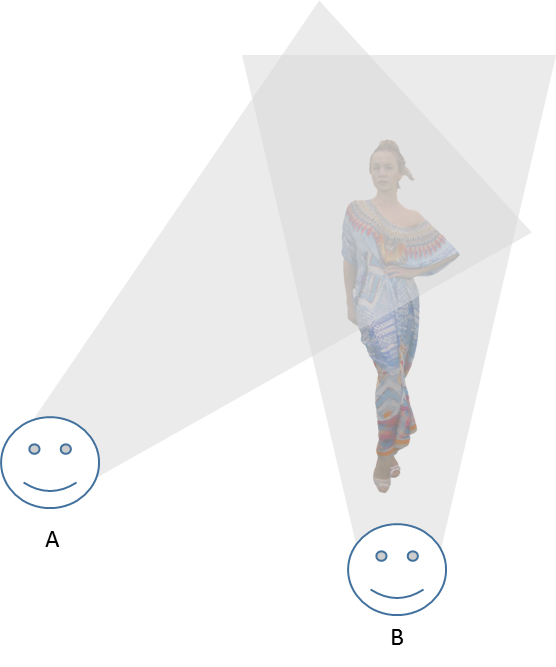
* Gathering of list of proposed solutions with overview descriptions: 2/7
* Gathering of use cases descriptions: 2/14
* Gathering evaluation criteria: 2/21
* Submission of evaluation results for discussion at AHG meeting: 3/13
* Submission of final evaluation results: 4/10
* Recommendation of a solution: during 130th MPEG meeting

# Description of CE

## Use case for evaluation

### Partial access of point cloud data for efficient access

An application represents a point cloud object[[1]](#footnote-1) as shown in Figure 1. Users have 6 DoF to view the object. Therefore, at any given point in time, a part of the available point cloud object will be visible to each of the users depending on the user’s location, viewport, field of view, etc.



**Figure 1 – Two users viewing subset of a point cloud object**

* User A will see upper right part of a point cloud object from left side of the object
* User B will see full of front part of a point cloud object from front of the object

In order to support the partial access and delivery of the point cloud object, there is a need to support for identifying one or more 3D spatial sub-regions to fully cover the user’s 6DoF viewport[[2]](#footnote-2) by the client, if necessary.

### Annotation[[3]](#footnote-3)

It is highly desirable to be able to annotate each region of the volumetric bitstream, i.e. the patches or groups of patches that are identified using a "rectangular" shaped volumetric rectangle, with different information. This process may include whether these elements are associated with a particular object (likely an object in the physical/world space) and certain properties that could be useful for their extraction and rendering. Such information may include labelling of objects, the size and shape of the points that correspond to the object, whether the object is visible or not, visibility cone information, material ids, and collision information, among others.

Ideally, we may wish to associate each such object with different parameters, or properties, which may also correspond to information provided during the creation or editing process of the point cloud, scene graph, etc. It is quite possible also that some objects may relate to one another and in some cases an object could be part of another object (e.g. we may have defined an object "man" or "woman", and other objects that correspond to "body", "head", "torso", "left hand", "right leg" etc).

An object could be persistent in time and could also be updated at any time/frame while the associated information may persist from that point onward. Multiple patches or 2D volumetric rectangles, which can contain themselves multiple patches, could be associated with a single object, but there may be a desire to also associate the same patch or 2D volumetric rectangles with multiple objects. Such relationships could persist or also need to change in time because objects may move or their placement in the atlas may have changed.

### Object-based scene synthesis[[4]](#footnote-4)

Introducing an identifier of a subset of points per patch enables novel use cases for volumetric video encoding, decoding, and rendering. In addition, 3D bounding box (may include label for a subset) per a subset can be signalled. This allows efficient identification, localization, labelling, tracking, and object-based processing. We list below few samples for innovative scene synthesis:

* Priority Objects Rendering:

With the identifiers for the subsets available at the decoder side, the renderer can select which subsets to output first (e.g. it may start rendering the subsets in the front first) while the subsets background / static subsets (may be inferred from the labels of the subsets within the associated SEI message) can be carried from last rendered intra-frame (to save compute and bandwidth). This helps speeding up processing at the decoding side to meet real-time requirements.

* Objects Filtering:

The user may choose to render only the subsets of interest while blur or remove other subsets. Also, the encoder may choose to pack patches only from those subsets of interest and transmit them rather than sending all in the event of limited bandwidth or probably may dedicate more bits to these patches to deliver the subsets of interests in higher resolution.

* Background Rendering:

Background is a special static subset that can be rendered by itself from the related patches or synthesized from virtual/ prerendered content. In case of rendering from Patches, there might be regions in the background that are not visible in any of the input source views due to occlusions. Such hole regions can be filled using inpainting techniques. Another approach is to capture the scene ahead without any subsets and stream within the metadata a single image once per intra-period so it can be used for rendering the background content and populate the scene with the subsets of interest afterward. A synthetic background can be inserted as well, and subsets can be augmented in.

## Evaluation criteria

1. Supporting flexible signalling of sub-divisions of a point cloud object into a number of 3D spatial sub-regions. Proponent can demonstrate flexibility of technology in terms of following point of view.
   * support of canonical sub-division (e.g. octree)
   * support of arbitrary sub-division
   * support of multi-level sub-division
   * etc.
2. Supporting efficient signalling of mapping between a 3D spatial sub-region of a point cloud object or a collection of them and one or more independently decodable subset of 2D video bitstream for partial access, delivery, and rendering,
3. Supporting the efficient rendering of a point cloud object across the time in a 6DOF environment.
4. Supporting viewport-adaptive delivery harmonization on definition of a 6DoF viewport,
5. Supporting a wide variety of use cases with the same syntax and semantics

# Related Contributions

## Apple

* m52705
* m52755

## Intel

* M52441
* M52442
* M52556
* M50654
* M50655

## InterDigital Communications, Inc.

* M47900
* M49107
* M49109
* M52257
* M52258

## Kyunghee University

## LG Electronics Inc.

* M47831
* M52285
* M52286
* M52288

## MediaTek

* M49868
* M49351
* M49352
* M49357
* M52493
* M52494

## Nokia

## Samsung

* M46040
* M47345
* M50772

## Sharp

## Sony

## Tencent

## ZTE

* M50807 Volumetric Tiling Information of Point Cloud Data

1. This term will be defined during the CE [↑](#footnote-ref-1)
2. This term will be defined during the CE [↑](#footnote-ref-2)
3. This use case has been extracted from the contribution m52705 [↑](#footnote-ref-3)
4. This use case has been extracted from the contribution m50950 [↑](#footnote-ref-4)