3GPP TSG SA WG4 Meeting #121S4-221363

14th – 18th November 2022

**Source:** Samsung Electronics Co., Ltd.

**Title: [MeCAR] Additional media types**

**Document for** Discussion and Agreement

# Introduction

At the last 120th meeting, the Media type categories and their characteristics were collected under section 3.7 of MeCAR PD v3.0 [1].

The section covers various AR/MR media types such as 3D content types (Scene description, 3D visual model) and metadata (spatial description, AR anchor, user pose, projection information, etc.) then identifies their real-time characteristics and traffic characteristics.

This contribution proposes to revise the tables in the section 3.7 with the followings.

* Improve text in the projection information
* Add split rendered media, in particular the 360-projection type in addition to the AR/MR media types
* Add depth map auxiliary video
* Add alpha auxiliary video

## Split rendered 2D projection video

The projection type of rendered media can be considered as planar perspective, orthogonal or omnidirectional projections. The definition and description column for the Projection information are proposed to improve text and to distinguish the information from media. In particular, signaling of omnidirectional projections from various specifications are listed as the references.

The Split rendered 2D projection video proposed in this contribution is a 2D video rendered and projected from 3D scene at the estimated pose. The projection type of the 2D video should be identified as it is relevant with the post-processes in the UE such as omnidirectional to planar projection conversion, pose correction or UE display specific projection. AVC, HEVC and OMAF can be used as the Split rendered 2D projection video with the projection type information referenced in the Projection information.

The Split rendered 2D projection video is also proposed as a primary picture for auxiliary video proposed in clause 1.2 and 1.3.

|  |  |  |
| --- | --- | --- |
| **AR/MR Media Type** | **Definition** | **Media Type Description (Examples)** |
| **Projection information** | Parameters associated to the projection to the 3D scene. Perspective, orthogonal and omnidirectional are the projection types to be used. | **Type**: In the case of omnidirectional projection, the signaling of projection type such as Equirectangular and Cubemap are defined as SEI in clause D.1.35 of ISO/IEC 14496-10 [X] and clause D.2.41 of ISO/IEC 23008-2 [Y], or as ISOBMFF box in clause 7.6.2 of ISO/IEC 23090-2 [Z]. |
| **Split rendered 2D projection video** | 2D video containing projected and rendered AR/MR media with the projection information. | **Type**: 2D video with projection information metadata.  **Organization:** MPEG |

|  |  |
| --- | --- |
| **AR/MR Media Type** | **Real Time Characteristics** |
| **Projection information** | Sparsely timed (event based) |
| **Split rendered 2D projection video** | Continuous |

|  |  |
| --- | --- |
| **AR/MR Media Type** | **Typical size or bitrate** |
| **Projection information** | TBD |
| **Split rendered 2D projection video** | TBD |

## Split rendered 2D depth map video

For UE to perform geometrically accurate pose correction, an auxiliary data which provides geometric understanding to the pose corrector may be used. Depth map is one of candidate information for the purpose. The depth representation types are defined in H.9.2.3 of AVC and G.14.3.3 of HEVC.

|  |  |  |
| --- | --- | --- |
| **AR/MR Media Type** | **Definition** | **Media Type Description (Examples)** |
| **Split rendered 2D depth map video** | Monochrome 2D auxiliary video containing depth map in the same perspective view with primary picture. | **Type:** 2D auxiliary video with depth type defined in I.13.2.3 of ISO/IEC 14496-10 [X] and G.14.3.3 of ISO/IEC 23008-2 [Y].  **Organization:** MPEG |

|  |  |
| --- | --- |
| **AR/MR Media Type** | **Real Time Characteristics** |
| **Split rendered 2D depth map video** | Continuous |

|  |  |
| --- | --- |
| **AR/MR Media Type** | **Typical size or bitrate** |
| **Split rendered 2D depth map video** | TBD |

## Split rendered 2D alpha channel video

To overlay virtual objects on the real world, transparency information should be provided with the split rendered video as described in section 3.6 of MeCAR PD. The carriage of transparency information is defined in 7.3.2.1.2 of AVC and F.3.5 of HEVC.

|  |  |  |
| --- | --- | --- |
| **AR/MR Media Type** | **Definition** | **Media Type Description (Examples)** |
| **Split rendered 2D alpha channel** **video** | 2D auxiliary video containing alpha channel information in the same perspective view with primary picture. | **Type:** 2D auxiliary video with alpha channel information defined in 7.3.2.1.2 of ISO/IEC 14496-10 [X] and F.14.3.8 of ISO/IEC 23008-2 [Y].  **Organization:** MPEG |

|  |  |
| --- | --- |
| **AR/MR Media Type** | **Real Time Characteristics** |
| **Split rendered 2D alpha channel video** | Continuous |

|  |  |
| --- | --- |
| **AR/MR Media Type** | **Typical size or bitrate** |
| **Split rendered 2D alpha channel video** | TBD |

# Proposed changes

## \*\*\* Start change 1 \*\*\*

### 3.7.2 Media types definition

Table 2 - AR/MR Media Type definitions

|  |  |  |
| --- | --- | --- |
| **AR/MR**  **Media Type** | **Definition** | **Media Type Description (Examples)** |
| **Scene Description** | Clause 4.4.2 of 3GPP TR 26.998[1]  Scene description is used to describe the composition of a 3D scene, referencing and positioning the different 2D and 3D assets in the scene typically using a tree or a graph structure. | **Type**: glTF2.0, JSON format  **Organization**: MPEG-I and Khronos |
| **Spatial Description** | Clause 4.4.7.3 of 3GPP TR 26.998[1]  Visual features, keyframes, and spatial maps are used for mapping the real world, typically as part of the SLAM process. | **Type**: For example:  <https://www.khronos.org/registry/OpenXR/specs/1.0/html/xrspec.html#XR_FB_spatial_entity>  **Organization:** OpenXR |
| **3D Visual Model** | Clause 4.6.3.5.2 of 3GPP TR 26.928[2]  3D visual object description as a list of vertices, faces and other elements, along with associated attributes. | **Type**: PoLYgon  **Organization:** None |
| **AR Anchor** | The AR anchor is meant to identify a point in the user space to be used to anchoring a visual object (2D or 3D) | **Type**: Metadata allowing accurate overlaying/rendering of text, graphics or video contents to support Use Case 8 of TR 26.928.  **Organization:** None |
| **User Pose** | Clause 4.4.3.1 of 3GPP TR 26.998[1]  Representation of the user position and orientation | **Type**: It consists of a quaternion for orientation and a 3D vector for position. Timestamp is represented by a 64 bit monotonically increasing nano-second-based integer.  **Organization**: Khronos OpenXR |
| **FOV** | Y.6.2.3 of 3GPP TS 26.114[3]  The Field of View (FOV) is the extent of observable world at any given moment | **Type**: It consists of vertical fov and horizontal fov.  **Organization:** None |
| **Viewport** | Y.7.2 of 3GPP TS 26.114[3]  The viewport corresponds to the projection of the user View onto a target display | **Type**: It shall contain all of the parameters Viewport\_azimuth, Viewport\_elevation, Viewport\_tilt, Viewport\_azimuth\_range and Viewport\_elevation\_range  **Organization:** None |
| **Gesture** | TBD | **Type**: A array of finger joint position.  For example: <https://www.khronos.org/registry/OpenXR/specs/1.0/html/xrspec.html#XR_EXT_hand_tracking>  **Organization:** OpenXR |
| **Body action** | TBD | **Type**: bvh format.  **Frequency**: at least 1kHz  **Organization:** BioVision company |
| **Facial expression** | OpenXR | **Type**: An array of key point position.  For example: <https://www.khronos.org/registry/OpenXR/specs/1.0/html/xrspec.html#XrSystemFacialTrackingPropertiesHTC>  **Organization:** None |
| **Sensor information** | TBD | **Type**: TBD  **Organization:** None |
| **Camera information** | TBD | **Type**: The camera parameters such as focal length, principal points, calibration parameters and the pose of the camera all contribute in understanding the relevance between points in the volumetric scene and pixels in the captured image.  **Organization:** None |
| **Projection information** | Parameters associated to the projection to the 3D scene. Perspective, orthogonal and omnidirectional are the projection types to be used. | **Type**: In the case of omnidirectional projection, the signaling of projection type such as Equirectangular and Cubemap are defined as SEI in clause D.1.35 of ISO/IEC 14496-10 [X] and clause D.2.41 of ISO/IEC 23008-2 [Y], or as ISOBMFF box in clause 7.6.2 of ISO/IEC 23090-2 [Z]. |
| **Split rendered 2D projection video** | 2D video containing projected and rendered AR/MR media with the projection information. | **Type**: 2D video with projection information metadata.  **Organization:** MPEG |
| **Split rendered 2D depth map video** | Monochrome 2D auxiliary video containing depth map in the same perspective view with primary picture. | **Type:** 2D auxiliary video with depth type defined in I.13.2.3 of ISO/IEC 14496-10 [X] and G.14.3.3 of ISO/IEC 23008-2 [Y].  **Organization:** MPEG |
| **Split rendered 2D alpha channel** **video** | 2D auxiliary video containing alpha channel information in the same perspective view with primary picture. | **Type:** 2D auxiliary video with alpha channel information defined in 7.3.2.1.2 of ISO/IEC 14496-10 [X] and F.14.3.8 of ISO/IEC 23008-2 [Y].  **Organization:** MPEG |

## \*\*\* End change 1 \*\*\*

## \*\*\* Start change 2 \*\*\*

Table 3 – Real Time Characteristics of AR/MR Media Types

|  |  |
| --- | --- |
| **AR/MR**  **Media Type** | **Real Time Characteristics** |
| **Scene Description** | Non timed |
| **Spatial Description** | Sparsely timed (event based) |
| **3D Model** | Non timed |
| **AR Anchor** | Sparsely timed (event based) |
| **User Pose** | Continuous |
| **FOV** | Non timed |
| **Viewport** | Continuous |
| **Gesture** | Continuous |
| **Body action** | Continuous |
| **Facial expression** | Continuous |
| **Sensor information** | Non timed |
| **Camera information** | Sparsely timed (event based) |
| **Projection information** | Sparsely timed (event based) |
| **Split rendered 2D projection video** | Continuous |
| **Split rendered 2D depth map video** | Continuous |
| **Split rendered 2D alpha channel video** | Continuous |

## \*\*\* End change 2 \*\*\*

## \*\*\* Start change 3 \*\*\*

Table 4 – Typical size and bitrate of AR/MR Media Types

|  |  |
| --- | --- |
| **AR/MR**  **Media Type** | **Typical size or bitrate** |
| **Scene Description** | TBD |
| **Spatial Description** | TBD |
| **3D Model** | TBD |
| **AR Anchor** | TBD |
| **User Pose** | TBD |
| **FOV** | TBD |
| **Viewport** | TBD |
| **Gesture** | TBD |
| **Body action** | TBD |
| **Facial expression** | TBD |
| **Sensor information** | TBD |
| **Camera information** | TBD |
| **Projection information** | TBD |
| **Split rendered 2D 360 video** | TBD |
| **Split rendered 2D depth map video** | TBD |
| **Split rendered 2D alpha channel video** | TBD |

## \*\*\* End change 3 \*\*\*

# Proposed

This contribution requests the proposed changes in section 3 to MeCAR PD v3.1.

# Reference

1. 3GPP S4-221272: “MeCAR Permanent Document v3.1”
2. 3GPP TR 26.998: “5G Glass-type Augmented Reality / Mixed Reality (AR/MR) devices”
3. ISO/IEC 14496-10: “Information technology — Coding of audio-visual objects — Part 10: Advanced video coding”
4. ISO/IEC 23008-2: “Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding”
5. ISO/IEC 23090-2: “Coded representation of immersive media - Part2: Omnidirectional media format”.