**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N18355**

**March 2019, Geneva, CH**

**Title: MPEG-I Phase 2 Use Cases**

**Source: WG11**

**Status: Public Draft**

**Editor: Eric Yip (eric.yip@samsung.com)**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Contents**

[1 Introduction 2](#_Toc488422708)

[2 Use Cases 2](#_Toc488422709)

# Introduction

This document gathers use cases for immersive media. They are used to derive requirements for immersive applications and services.

Use cases presented here target the second phases of MPEG-I : Phase 2.

A separate document, MPEG-I Phase 1 Use Cases (N17504), already exists to specifically gather use cases for MPEG-I Phase 1a and 1b.

# Use Cases

|  |
| --- |
| Combined Point Cloud and Video 6DoF Contents |
| Description  A user is watching a sports match, or a concert, using a device with the capability to provide input from the user to enable him/her to change his/her viewpoint location and direction within the sport or concert venue, without restriction.  Image result for FreeD intelImage result for FreeD intelFor example, the user has the possibility to select a viewpoint from a 1st person perspective of a sports player, or a 3rd person perspective viewpoint similar to that of more traditional TV broadcast sports contents (figure 1).    Figure 1: an example of 2 different viewpoints which could be selected by a user. Left: 1st person player view. Right: 3rd person commentary view [2].  The 6DoF content which the user is viewing is rendered using a combination of both point cloud media data, and video media data. The whole sports or concert venue is captured using multiple high resolution cameras, such that the video data captured can be processed to create a point cloud scene of the centre of the sports or concert venue (e.g. the sports pitch, players and other dynamic objects are represented by point clouds in the scene). This processing can be performed either at the venue itself, or remotely on a dedicated network.  By creating such point cloud media data, a user has the freedom to navigate within the sports or concert venue (i.e. the defined scene boundary here), and is able to view different players and objects from all viewpoints and positions.  Since such venues are traditionally very big, and include massive crowds of spectators, it is possible to represent such non-interactive parts of the scene background using video media data.  The result is that the user views both point cloud and video media rendered at the same time in order to create an immersive experience.  Further applications of this use case include the smooth changing of views: from views containing only video media data, to combined point cloud/video media data, and to views containing only point cloud media data. Such smooth changing of views is especially important for rendering devices which allow restricted viewpoint changes, or for playback applications such as some predefined guided viewing. Sports and concert events typically have multiple professional broadcasting cameras capturing the event at different locations with different panning trajectories; by supporting the smooth transition of views from video to combined media, both the high quality and artistic intent of video captured by broadcasting cameras can still be portrayed to the user as part of the immersive experience.  Since point cloud media data of a complete sports or concert scene typically takes up huge amounts of data, the processing, delivery and rendering of media in this use case should be dependent on the view selected or defined by the user. Through the user selected viewpoint location and direction, relevant media data should be sent accordingly to the user; functionalities specific to point cloud navigation, such as level of density adaption, should be considered in order to achieve low latency in the system. |
| **Overlap with other use cases** |
|  |
| **Target Phase (1a, 1b, 2) :** Phase 2 |
| **Required  features** |
|  |
| **Identified Gaps and Optimization Potentials** |
| **Potential Requirements and Specifications** |
| The MPEG-I 6DoF system may provide functionalities to support this use case through the following key requirements:   * The system shall support the processing (for media distribution),of different media types. * The system shall support the storage of different media types. * The system shall support the presentation and playback of different media types. * *Note: Media types may include,2D video, spherical video, point cloud and various audio media data* * The system shall support the processing and storage, presentation of multiple different media types such that they are synchronized both spatially and temporally for simultaneous rendering, more specifically a 2D background video surrounding a point cloud object. * The system shall support low-delay delivery and rendering of the above media types for 6DoF contents~~.~~ * The system shall support metadata which allows natural and smooth switching between real camera captured views and other virtually processed camera views within the point cloud scene. * The system shall support the partial delivery and rendering of content media depending on the user selected location and view within the point cloud scene. * The system shall support level of density adaptation of point cloud media depending on the user selected view location and end-to-end system capabilities. * The system may support the processing of multicamera media data to create the point cloud scene. |

|  |
| --- |
| Scene and object feature based 6DoF media change |
| A user is watching a sports match produced in 6DoF on a mobile device which is capable of rendering 6DoF media content, but is only available to capture 2D media by its camera module.  In the middle of the match, there are some moments to replay the important shots of a player. When the player is zoomed in and the camera moves along the player, changing the player’s face to the user’s face can be provided as an entertainment purpose.  For example, the player’s crying face can be replaced by user’s face which is transformed from the 2D image or short video taken by the camera installed in user’s mobile device.  Regardless the expression of user’s face or skin colors and textures, and regardless the lower DoF of user’s face data, the user’s face is naturally absorbed into the 6DoF player’s face so that user’s crying face is shown with the player’s body. |
| **Overlap with other use cases** |
|  |
| **Target Phase (1a, 1b, 2) :** Phase 2 |
| **Required  features** |
|  |
| **Identified Gaps and Optimization Potentials** |
| **Potential Requirements and Specifications** |
| The MPEG-I 6DoF system may provide functionalities to support this use case through the following key requirements:   * The system shall support metadata which includes information describing objects and object features within a scene. * The system shall support metadata for processing of object features, namely preserving and replacing features for objects in a scene. * The system shall support construction of 6DoF media data from lower degree of freedom media data. * The system shall support low delay processing of object feature analysis. * The system shall support low delay processing and rendering of object features for a scene. |

|  |
| --- |
| Remote pre-rendering for 6DoF contents |
| This use case reflects the full immersive content experience which will be enabled through live captured content. When a live sports game, such as Super Bowl, is offered as 6DoF immersive content, a user would receive a 6DoF volumetric video and watch the game from a perspective of his favourite player in the game. In some cases such as a mobile terminal, content are sent to a remote renderer at the mobile edge clouds for pre-rendering according to a user’s selected viewport, the network conditions, and/or the device capabilities and delivered to a client terminal. |
| **Overlap with other use cases** |
| **Target Phase (1a, 1b, 2) :** phase 2 |
| **Required  features** |
|  |
| **Identified Gaps and Optimization Potentials** |
| **Potential Requirements and Specifications** |
| 1. The MPEG-I system shall support full or partial delivery media content depending on different network conditions and device capabilities and configurations. 2. The system shall support (pre-)rendering of immersive contents by a clouds system instead of the client which will consume the contents, where the output will be either 6DoF, 3DoF+, 3DoF, or 2D. 3. The system shall support remote encoding for converting immersive 6DoF content to a simpler representation such as 2D image/video. 4. The system shall support 6 DoF media contents metadata for (pre-)rendering by a clouds system. 5. The system shall support user metadata (different network conditions and device capabilities and configurations) for (pre-)rendering by a clouds system. 6. The system shall support the network-based media processing framework |

|  |
| --- |
| Full Immersive Content: 6DoF with full 3D 360 video |
| Description  This use case reflects the full immersive content experience which will be capable through real-life captured content in the future (defined by MPEG-I). Full 6DoF immersive content allows the viewer to navigate his/her location within the content space, freely without limitation, with natural change of the corresponding rendered video and audio. The media data will involve new data concepts (such as light fields), most likely requiring needs for new capturing technologies, video and audio codecs, delivery systems as well as new display technologies. The result of using these new technologies will provide a high quality, realistic immersive experience.    Figure 2: an example of full 6DoF immersive content where a viewer may change his or her viewpoint to any location within the scene boundary |
| **Overlap with other use cases** |
|  |
| **Target Phase (1a, 1b, 2) :** Phase 2 |
| **Required  features** |
| Full 6DoF, Full 3D |
| **Identified Gaps and Optimization Potentials** |
| **Potential Requirements and Specifications** |
|  |

|  |
| --- |
| A mobile Subject looks around a statue Object |
| A single person moves around a still statue object in a room and looks at the statue. |
| **Overlap with other use cases** |
|  |
| **Target Phase (1a, 1b, 2) :** 2 |
| **Required  features** |
| The media content is of sphere type, the object is inside the sphere, and the subject looks at the object from outside of the sphere, and has the 3 rotational DoFs plus 2 translational DoFs (no movement in the z-axis). |
| **Identified Gaps and Optimization Potentials** |
|  |
| **Potential Requirements and Specifications** |
| Media Format shall support   * a list of media geometry types, including sphere and rectangular box * switching among content with different geometry types * how content at different locations are spatially related   Presentation Format shall support   * viewports with translations and orientations in 3/2/1 dimensions * viewport dependent presentation that can be adapted to different network conditions and device capabilities and configurations   Orchastration Format shall support   * orchestartion in time * orchestartion across space |

|  |
| --- |
| A mobile Subject looks around a statue Object with up-close views |
| A single person moves around a still statue object in a room and looks at the statue, with the capability to look closely. |
| **Overlap with other use cases** |
|  |
| **Target Phase (1a, 1b, 2) :** 2 |
| **Required  features** |
| The media content consists of multiple nested spheres with the same center, and the object is at the center of the spheres, the subject looks at the object from outside of the spheres, and has the 3 rotational DoFs plus 2 translational DoFs (no movement in the z-axis). The subject is capable of view changing from sphere to sphere. |
| **Identified Gaps and Optimization Potentials** |
|  |
| **Potential Requirements and Specifications** |
| Media Format shall support   * a list of media geometry types, including sphere and rectangular box * content with nested geometry type * switching among content with different geometry and nested geometry types. * how content at different locations are spatially related   Presentation Format shall support   * spherical viewports with translations and orientations in 3/2/1 dimensions, moving along the z-axis * viewport dependent presentation that can be adapted to different network conditions and device capabilities and configurations   Orchastration Format shall support   * orchestartion in time * orchestartion of the nested spherical content along the z-axis |

|  |
| --- |
| Multiple subjects look around in a Room with a tour guide |
| A group of people stand still in a center of a room of the museum and looks around, audio guided by a tour guide. |
| **Overlap with other use cases** |
|  |
| **Target Phase (1a, 1b, 2) :** 2 |
| **Required  features** |
| The media content is of the sphere type with audio, subjects are at the center of the sphere, looking from inside out, each with the 3 rotational DoFs, but their views can be synchronized by the guide audio – a social media experience. |
| **Identified Gaps and Optimization Potentials** |
|  |
| **Potential Requirements and Specifications** |
| Media Format shall support   * a list of media geometry types, including sphere and rectangular box * switching among content with different geometry and nested geometry types * how content at different locations are spatially related   Presentation Format shall support   * viewports with translations and orientations in 3/2/1 dimensions * viewport dependent presentation that can be adapted to different network conditions and device capabilities and configurations   Orchastration Format shall support   * orchestartion in time * orchestartion across space * orchestration in terms of logical compositions |

|  |
| --- |
| Multiple subjects look around a statue with a tour guide |
| A group of people move around a still statue object, all looking at the statue, with different viewpoints, but audio guided by a tour guide. |
| **Overlap with other use cases** |
|  |
| **Target Phase (1a, 1b, 2) :** 2 |
| **Required  features** |
| The media content is of sphere type, the object is inside the sphere, and multiple subjects look at the object from outside of the sphere, with potentially different viewports, but their views can be synchronized by the guide audio – a social media experience. |
| **Identified Gaps and Optimization Potentials** |
|  |
| **Potential Requirements and Specifications** |
| Media Format shall support   * a list of media geometry types, including sphere and rectangular box * content with nested geometry types * switching among content with different geometry and nested geometry types. * how content at different locations are spatially related   Presentation Format shall support   * viewports with translations and orientations in 3/2/1 dimensions * viewport dependent presentation that can be adapted to different network conditions and device capabilities and configurations   Orchastration Format shall support   * orchestartion in time * orchestartion across space * orchestration in terms of logical compositions |

|  |
| --- |
| Inter-change experiences |
| People can change their touring experiences between looking around within a room and looking at a statue. |
| **Overlap with other use cases** |
|  |
| **Target Phase (1a, 1b, 2) :** 2 |
| **Required  features** |
| Switching between 3DoF/3DoF+ navigations and converging navigations around an object |
| **Identified Gaps and Optimization Potentials** |
|  |
| **Potential Requirements and Specifications** |
| Media Format shall support  The media format is one used for describing captured and processed media content   * It shall support a list of media geometry types, including sphere and rectangular box * It shall support content with nested geometry types * It shall support switching among content with different geometry and nested geometry types * It shall support how content at different locations are spatially related   Presentation Format shall support  The presentation format is one used for presenting captured and processed media content on one or more display devices   * It shall support viewports with translations and orientations in 3/2/1 dimensions * It shall support viewport dependent presentation that can be adapted to different network conditions and device capabilities and configurations   Orchastration Format shall support  The orchestration format is one used for organizing and orchestrate captured and processed media content potentially from one or more sources onto one or more display devices   * It shall support orchestartion in time * It shall support orchestartion across space * It shall support orchestration in terms of logical compositions |

|  |
| --- |
| VR Video Calling |
| 1. Panoramic video calling : Alex and Bob are at different physical locations, and they are having a video calling, Alex send 360 video to Bob while Bob send 2D video cause his devices with limited capabilities (cannot take 360 videos)  * Bob wears an HMD to watch the 360 video sent by Alex * If Bob doesn’t have an HMD, he can only play the 360 video on other terminals (mobile phone, PC, etc.)     (https://www.insta360.com/)   1. Alex and Bob in the same VR environment while Bob send 2D video  * Alex can be present in the VR environment through some form of user-embodiment while Bob’s 2D video will be present as a VR object (virtual screen, etc.) in the VR environment, which Alex can interact with (move, zoom in/out, etc.)   C:\Users\cmri\AppData\Local\Temp\1553443447(1).png  (https://www.facebook.com/spaces) |
| **Overlap with other use cases** |
| * Multiple users in VR environment * Social TV and VR * Multiple users in VR environment, 6DoF * VR Conferencing |
| **Target Phase (1a, 1b, 2) : Phase 1b, Phase 2.,** |
| **Required  features** |
| * Interactions with VR objects * Synchronization of audio and video of users and the scene * Users whose devices with limited capabilities (without motion tracking) can get into a shared VR environment |
| **Identified Gaps and Optimization Potentials** |
|  |
| **Potential Requirements and Specifications** |
| * Detection & rendering of user interactions with VR objects * User shall be present in the VR environment through some form of user-embodiment * Zooming in/out of region of interest by user interactions * User’s 2D video can be rendered as a VR object (virtual screen, etc.), which other users can interact with (move, zoom in/out, etc.) |