**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG/m52412**

**January 2019, Brussels, Belgium**

|  |  |
| --- | --- |
| **Source** | **Interdigital** |
| **Status** | **Contribution** |
| **Title** | **Simplified Viewing Space Specification** |
| **Author** | Renaud Doré, Bertrand Chupeau |

# Introduction

A syntax for carrying viewing space specification was adopted at MPEG 128 in the Working Draft 3 of the MIV standard [1]. It enables the description of complex shapes – including a viewing direction constraint – either based on CSG or interpolation operations which prove to be complementary approaches depending on the acquisition camera rig geometry [2].

However there are typical 3D content acquisition conditions for which a viewing space can be straightforwardly deduced from the camera rig geometry. This contribution proposes a simplified viewing space specification syntax depending upon extrinsic camera parameters, as a lightweight alternative to the current specification.

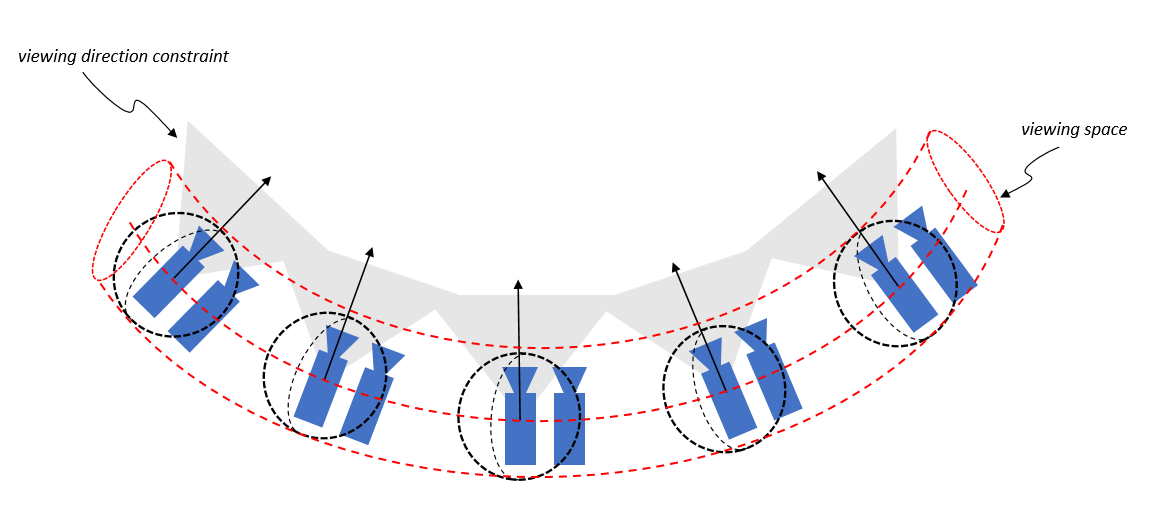


Figure 1: Specification of a tuboïd viewing space from five source cameras of PoznanFencing

# Rationale

PoznanFencing test sequence [3] is a good example of such 3D content the viewing space of which is straightforwardly inferred from the acquisition rig. Five stereo pairs of convergent cameras are arranged along a circular arc: a tuboïd-shaped viewing space can thus be defined by interpolation from a set of primitive shapes sampled along the arc.

If the sampled primitive shapes are positioned at camera locations, there is neither need to signal their position nor their viewing directions, but only the selected cameras. The example of PoznanFencing is depicted in Figure 1, with five spheroids positioned at every other camera location to fully specify the viewing space.

# Proposal

Our simplified alternative viewing space specification consists of:

* Selecting ordered subsets of cameras to define elementary shapes by interpolation
* Restricting to spheroid primitive shapes

As the extrinsic camera parameters, cam\_pos\_x[ v ], cam\_pos\_y[ v ], cam\_pos\_z[ v ], cam\_yaw[ v ], cam\_pitch[ v ] are already present in view\_params\_list( ), only the camera indices v need to be signaled, along with the spheroid radii and the yaw and pitch ranges.

# Viewing space specification update

## Syntax

|  |  |
| --- | --- |
| viewing\_space( ) { | **Descriptor** |
| **num\_elementary\_shapes\_minus1** | u(v) |
| **camera\_inferred\_flag** | u(1) |
| for( e = 0; e <= num\_elementary\_shapes\_minus1; e++) { |  |
| **elementary\_shape\_operation**[ e ] | u(1) |
| if( camera\_inferred\_flag == 0 ) |  |
| elementary\_shape( e ) |  |
| else |  |
| camera\_inferred\_elementary\_shape( e ) |  |
| } |  |
| } |  |

|  |  |
| --- | --- |
| camera\_inferred\_elementary\_shape( e ) { | **Descriptor** |
| **num\_spheroids\_minus\_1**[ e ] | u(v) |
| **guard\_band\_present\_flag**[ e ] | u(1) |
| **viewing\_direction\_constraint\_present\_flag**[ e ] | u(1) |
| for( s= 0; s <= num\_spheroids\_minus1[ e ]; s++ ) { |  |
| **view\_idx**[ e ][ s ] | u(v) |
| **radius\_x**[ e ][ s ] | fl(16) |
| **radius\_y**[ e ][ s ] | fl(16) |
| **radius\_z**[ e ][ s ] | fl(16) |
| if(guard\_band\_present\_flag[ e ] ) |  |
| **guard\_band\_size**[ e ][ s ] | fl(16) |

|  |  |
| --- | --- |
| if( viewing\_direction\_constraint\_present\_flag[ e ] ) { |  |
| if( guard\_band\_present\_flag[ e ]) |  |
| **guard\_band\_direction\_size**[ e ][ s ] | fl(16) |
| **primitive\_shape\_viewing\_direction\_yaw\_range**[ e ][ s ] | fl(16) |
| **primitive\_shape\_viewing\_direction\_pitch\_range**[ e ][ s ] | fl(16) |
| } |  |
| } |  |
| } |  |

## Semantics

**camera\_inferred\_flag** equal to 1 indicates that a lightweight elementary shape specification depending on camera parameters is used. camera\_inferred\_flag equal to 0 specifies that the unconstrained elementary\_shape( ) specification is used.

**num\_spheroids\_minus\_1**[ e ] specifies the number of spheroids which define (by interpolation operation) the e-th elementary shape; num\_spheroids\_minus\_1[ e ] shall be in the range 0 .. num\_views\_minus1.

**view\_idx**[ e ][ s ] specifies the index of the camera the location of which the s-th spheroid of the e-th elementary shape is placed at.

**radius\_x**[ e ][ s ], **radius\_y**[ e ][ s ], **radius\_z**[ e ][  s ] are 16-bit floating-point values that specify the dimension x, y and z respectively of the s-th spheroid in the scene coordinate system.

# Conclusions and recommendations

We recommend to adopt the proposed syntax and semantics modifications for the viewing space specification.

# References

1. “Working Draft 3 of Metadata for Immersive Video,” MPEG document N18794, ISO/IEC JTC 1/SC 29/WG 11, Geneva, CH, October 2019.
2. “Interdigital-Nokia response to Immersive Video CE-1 (Viewing Space)”, MPEG document M51627, ISO/IEC JTC 1/SC 29/WG 11, Brussels, Belgium, January 2020.
3. “Common Test Conditions for Immersive Video,” MPEG document N18789, ISO/IEC JTC 1/SC 29/WG 11, Geneva, CH, October 2019.