 ISO/IEC JTC 1/SC 29/WG 7 N00112

**ISO/IEC JTC 1/SC 29/WG 7  
MPEG 3D Graphics Coding   
Convenorship: AFNOR (France)**

**Document type:** Output Document

**Title:** Use cases for Mesh Coding

**Status:** Approved

**Date of document:** 2021-05-07

**Source:** ISO/IEC JTC 1/SC 29/WG 7

# Expected action: None

# Action due date: None

**No. of pages:** 6 (with cover page)

**Email of Convenor:** marius.preda @ imt . fr

**Committee URL:** [https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg7](https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3)

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 7 MPEG 3D Graphics Coding**

**ISO/IEC JTC 1/SC 29/WG 7 N** **00112**

**April 2021, Virtual**

|  |  |
| --- | --- |
| **Title** | Use cases for Mesh Coding |
| **Source** | **WG 7, MPEG 3D Graphics Coding** |
| **Status** | **Approved** |
| **Serial Number** | **20364** |

# Introduction

While so far the most common way of representing the visual component of the world has been to take the output of a camera, compress it for transmission and storage using one of the MPEG video coding standards and eventually decode it and present it on 2D displays, there are now more and more devices that capture and present 3D representations of the world.

A representation of 3D objects or scenes can be a 3D mesh, which is composed of a set of vertices of connected polygons and with additional connectivity information. The 3D mesh may also carry attributes such as color, normal, reflectance associated with each vertex. It may alternatively use texture maps to carry attribute(s) information associate with the surface of the polygons. In that case, UV coordinates are associated with each vertex, mapping the 3D model’s surface onto a 2D image. Besides color information, texture mapping can also be used to carry additional information, such as height mapping, bump mapping, normal mapping, displacement mapping, reflection mapping, specular mapping, occlusion mapping, and many others.

As compression technologies are needed to reduce the amount of data required to represent a mesh and as there is no MPEG standard to address dynamic meshes (with time varying attributes and connectivity information), MPEG is planning to develop compression standards for volumetric content represented by dynamic meshes targeting lossless, near-lossless and lossy compression (for use in real-time communications), with attributes of efficient geometry and attributes compression, scalable/progressive coding, coding of sequences of meshes captured over time, and random access to subsets of the meshes.

The acquisition of meshes is outside of the scope of this standard.

# Use Cases for Mesh Compression

This document presents the use cases targeted for the MPEG standard on mesh compression. Each of the use cases are used to illustrate the requirements defined in [1] and [11].

## Real-Time 3D immersive telepresence

Advances in 3D capture and reconstruction enable real-time generation of highly realistic 3D representations for 3D tele-presence. 3D meshes are an efficient representation as they can be seamlessly integrated and rendered in 3D virtual worlds enabling a convergence between real and virtual realities. In this case, meshes are reconstructed, compressed transmitted and rendered in real-time as in video conferencing systems, enabling conversational style communication.

3D Reconstruction

Software

Reconstructed 3D Human

3D Mesh

Representation

3D Source Encoding

IP

Network

Multi-depth camera

capture

or other 3D capture

Real-Time 3D Rendering

Composition

In Virtual World

Packetization

&

Transmission

3D Source Decoding

N RGB + Depth Images

Or other sensor data



Reception

&

Synchronization



Figure 1: Transmission pipeline for conferencing with 3D geometry

Figure 1 shows a typical tele-immersive media pipeline. Multiple calibrated color plus depth cameras are used to acquire streams that are fed into a reconstruction module (sensed data converter [2]). The reconstruction module produces 3D meshes based on a reconstruction algorithm. Reconstructed meshes are then compressed and transmitted and at the receiver side decoded and rendered compositely in a virtual world or 3D space. The 3D Reconstruction Software could split an object into parts and encode them as mesh in order to achieve optimum visual quality. Some examples of these systems in the industry, 8i [3], Microsoft holo-portation [4], Holocap [8],and in ongoing industry-oriented research projects [5].

**Key requirements for this application are:**

1. *lossy compression* with bit-rate control is needed
2. *low complexity and/or support for real-time* encoding/decoding is needed
3. *error resilience* to transmission errors is needed
4. *texture maps* or *vertex attributes* coding is needed for realistic rendering
5. *material/normal properties* coding is needed to support the rendering
6. *view dependence:* for streaming optimization view dependence can be used to optimize the transmission process

**Typical meshes in this use case have the following characteristics:**

1. Between 40,000 and 100,000 triangles with color per vertex to represent a reconstructed human
2. Between 10,000 and 50,000 triangles with texture maps to represent a reconstructed human
3. Texture maps for color representation, between 2K square pixels to 8K square pixels
4. Normals and or material properties to support the rendering using a shader

## Content AR/VR viewing with Interactive Parallax

New AR/VR generation now provides to end-users equipped with Head Mounted Display (HMD) an immersive experience for viewing specific contents like large field of view movies. One major improvement for improving the visual comfort relates to the viewing of content with interactive parallax, where the rendering viewport is updated for each new position of the end-user head. This use case is very much related to the previous one, except that

* It does not require a real time preparation of the source but is rather typically based on a long and sophisticated production workflow mixing shooting and computer graphics. The production workflow can split the objects or scene into parts and encode them as meshes in order to achieve optimum visual quality.
* The meshes in some cases only needs to be rendered from a range of viewport position, typically corresponding to a bounding box around the average position of the end-user head. This corresponds to the view dependence mentioned here below.
* The meshes may be a part only of the total content transmitted to the rendering sites.

Content production workflow

Packetization

& Transmission

3D Source (non real time) encoding



IP network

Real time 3D rendering with HMD

and composition in virtual world

Reception & synchronization

3D Source (real time) decoding

Figure 2: Meshes for interactive parallax broadcasting

**Key requirements for this application are:**

1. *lossy compression* with bit-rate control is needed
2. *error resilience* to transmission errors is needed
3. *texture maps* or *vertex attributes* coding is needed for realistic rendering
4. *material/normal properties* coding is needed to support the rendering
5. *view dependence:* the view dependence can be used to optimize the encoding process

**Typical meshes in this use case have the following characteristics:**

1. Between 10,000 and 100,000 triangles to represent a reconstructed human
2. Texture maps for color representation, between 2K square pixels to 8K square pixels
3. Normals and or material properties to support the rendering using a shader
4. Global parameters defining the spatial constraints of the rendering viewport

## 3D Free viewpoint Sport Replays Broadcasting

Meshes capture of sports events like basketball, baseball for free viewpoint playback and interaction on mobile devices and TV requires a compression and file format standard such that the reconstructed meshes can be streamed or stored in interoperable manner. Desired features are good compression efficiency, progressive download for different device capabilities. The industry has already moved in this direction, examples are Replay Technology [7] (see Figure 3). The full scene can also contain textures and/or video. The coding of these data is outside the scope of this standard and can be performed using alternative MPEG standards. The scope of meshes coding will only be to compress the meshes. Composition and synchronization can be achieved using MPEG Systems standards.

|  |  |  |
| --- | --- | --- |
|  |  |  |

Figure 3 Free viewpoint sport capture as meshes

**Key requirements for this application are:**

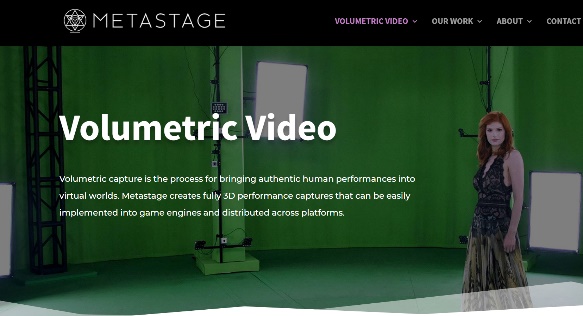
1. *low delay* encoding and decoding
2. *texture maps* or *vertex attributes* coding is needed, preferable 8-12 bits per component

**Typical meshes in this use case have the following characteristics:**

1. Between 20,000 and 200,000 triangles to represent reconstructed humans
2. Texture maps of 8-12 bits per color component
3. Can contain multiple clusters/groups of individual meshes (different players)

**Annex** - **Volumetric capture**

There are several emerging studio solutions providing new capabilities to capture and distribute 3D content. Companies like Holocap [8], Metastage [9], 4DViews [10], HHI [12], Volucap [13] are using dynamic mesh content with high number of polygons varying in each frame to represent a diverse set of objects, like animals and humans.

****

# References

[1] n16330 Requirements for Point Cloud Compression, ISO JCT1 SC29 WG11 Geneva June 2016

[2] m37050 3D Visual Capture Inventory ISO JCT1 SC29 WG11 Geneva October 2015

[3] “Making virtual Reality Human” Create Volumetric video of real people <http://8i.com/>

[4] “Microsoft Holoportation” <http://research.microsoft.com/en-us/projects/holoportation/>

[5] “Real and Virtual engagement in Realistic Virtual Environments” [www.reveriefp7.com](http://www.reveriefp7.com)

[6] P. van Oosterom, O. Martinez-Rubi, M. Ivanova, M. Horhammer, D. Geringer, S. Ravada, Theo Tijssen, Martin Kodde, Romulo Gonçalves, Massive point cloud data management: Design, implementation and execution of a point cloud benchmark, Computers & Graphics, Volume 49, June 2015, Pages 92-125, ISSN 0097-8493, http://dx.doi.org/10.1016/j.cag.2015.01.007

[7] Replay technologies <http://replay-technologies.com/>

[8] <https://holocap.com/>

[9] <https://metastage.com/>

[10] <https://www.4dviews.com/>

[11] MDS20363, Draft Requirements for Mesh Coding, ISO/JCT SC29 WG7 Online April 2021

[12] <https://www.hhi.fraunhofer.de/en/departments/vit/research-groups/immersive-media-communication/research-topics/volumetric-video-production.html#c39679>

[13] <https://volucap.de/>