INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 7
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC 1/SC 29/WG 7 **m 55363**
Online – October 2020

Title: **[VPCC] [Requirements] Update of requirements for V-PCC v2**

Author: Danillo Graziosi, Ohji Nakagami, Alexandre Zaghetto, Satoru Kuma, Ali Tabatabai, Sebastian Schwarz, Madhukar Budagavi, Rajan Joshi

1 Introduction

In this contribution, we list new use-cases that should be considered by the group for targeting mesh compression. The goal is to include in the requirements data types that are believed to be of interest from the industry and academia. The purpose of this contribution is to present new cases and start a discussion on either extending the current PCC project, or creation of a new part, based on the V3C specification. Additionally, we hope that this contribution act as a catalyst for technical contributions on mesh compression.

Formatted: Not Highlight

Formatted: Not Highlight

~~In this contribution, we list new use-cases that should be considered by the group for the next version of the V-PCC standard targeting mesh compression. The goal is to include in the requirements for V-PCC v2 data types that are believed to be of interest from the industry and academia.~~

~~3~~2 Proposal

Mesh content creation has had a remarkable evolution recently, especially with the advances in face/body capture and multi-view acquisition [1][2][3]. Furthermore, deep learning techniques also provide powerful tools for mesh content generation, creating realistic content with finer details and tracked features. Nevertheless, the meshes created do not follow the narrow scope of voxelized point clouds. Therefore, we propose to include in the requirements for mesh compression the following new use-cases:

- Floating-point and Voxelized mesh content
- Registered and Non-registered mesh content
- Rigged mesh content
- Mixed content
  - Tracked and untracked mesh content
    - (clean content)
  - Tracked/untracked mesh and point cloud content
    - (clean and noisy content)

Formatted: English (Canada)

43 Mesh formats for consideration

4.13.1Floating point meshes

It is very common to find meshes, where the vertex position is defined using floating points. For instance, Figure 1 shows a mesh downloaded from [4]

```
# 3ds Max Wavefront OBJ
# File Created: 06.12.2018 16:03:17
#
# object Jess_Casual_Walking_001_Vray
#
mtllib Jess_Casual_Walking_001.mtl
v  -44.4140 -107.0191 33.3839
v  39.3553 -319.0138 19.4809
v  76.1228 220.1020 113.6189
v  71.5332 143.1668 223.8207
v  20.6617 -191.5423 96.4831
v  49.3052 200.5206 282.1293
v  -58.2500 -108.6439 1363.9957
```

Figure 1 : Example of floating-point mesh

~~4.2~~3.2Registered (tracked) meshes

To improve compression and rendering, many solutions deploy mesh tracking, and several successful techniques have been published [5][6][7]. Furthermore, motion capture techniques are now able to add realism to computer generated models and is being widely deployed in gaming platforms [2]. Unlike dynamic, unregistered meshes, tracked meshes can be efficiently transmitted, since the topology does not change every frame. However, to add more realism, tracked meshes are only allowed a certain number of frames, and the texture might change with different lighting conditions. Even though tracked meshes are more compact, they still can generate a significant amount of data and require compression.

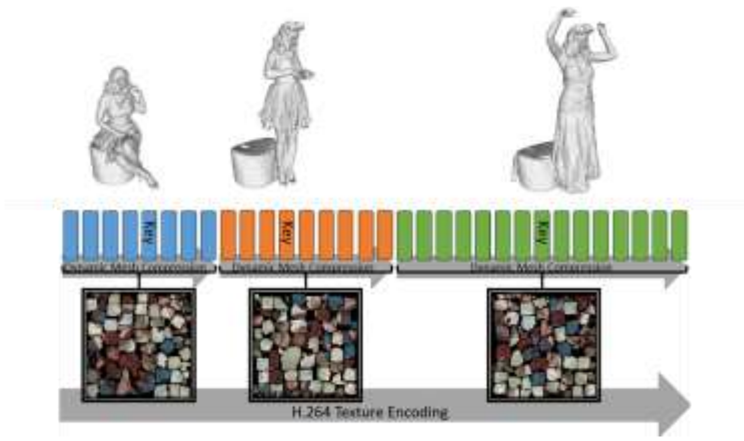

Figure 2: Tracked meshes, from [3]

3.3 Rigged mesh content

Rigging is a technique in computer animation in which a character (or other articulated object) is represented in two parts: a surface representation used to draw the character (i.e. the mesh) and a hierarchical set of interconnected parts (called bones, and collectively forming the skeleton or rig), a virtual armature used to animate (pose and keyframe) the mesh [9]. When the animated object is more general than, for example, a humanoid character, the set of "bones" may not be hierarchical or interconnected, but simply represent a higher-level description of the motion of the part of mesh it is influencing. This technique is often used to animate humans and other organic figures.

The process of rigging is constructing a series of bones (which need not correspond to any real-world anatomical feature). Each bone has a three-dimensional transformation from the default bind pose (which includes its position, scale and orientation), and an optional parent bone. The bones therefore form a hierarchy. The full transform of a child node is the product of its parent transform and its own transform. So, moving a thigh-bone will move the lower leg too. As the character is animated, the bones change their transformation over time, under the influence of some animation controller. A rig is generally composed of both forward kinematics and inverse kinematics parts that may interact with each other. *Skeletal animation* is referring to the forward kinematics part of the rig, where a complete set of bone configurations identifies a *unique pose*.

For encoding of such animated content V3C should develop efficient representations of the *unique pose* of the rig, as well as its *skeletal animation*.

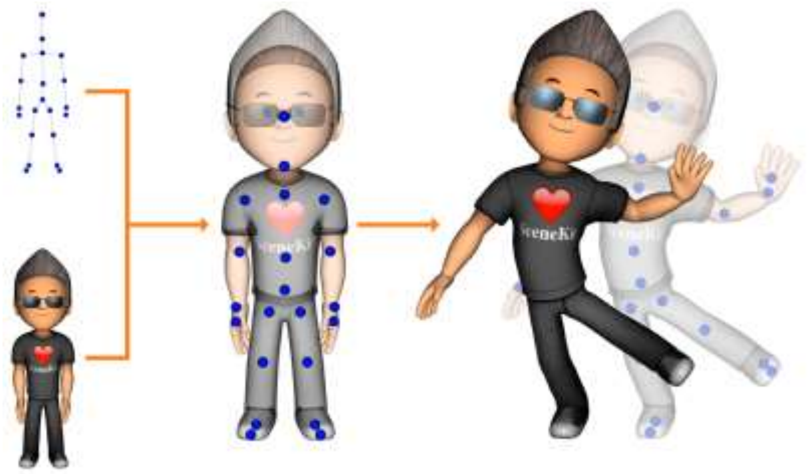

Figure 3: Example of rigging (middle) a skeleton (top) to a mesh (bottom) for skeletal animation (right). Source: https://developer.apple.com/documentation/scenekit/scnskinner

4.33.4 Mixed meshes

One way to deal with changes in topology is to reset the entire frame, as was done in [3]. However, a significant part of the mesh could still be tracked, since not all parts of the mesh are affected by topology changes in the same way. For instance, in [8], the authors use UV maps and mix tracked with untracked content, as shown in Figure 4.

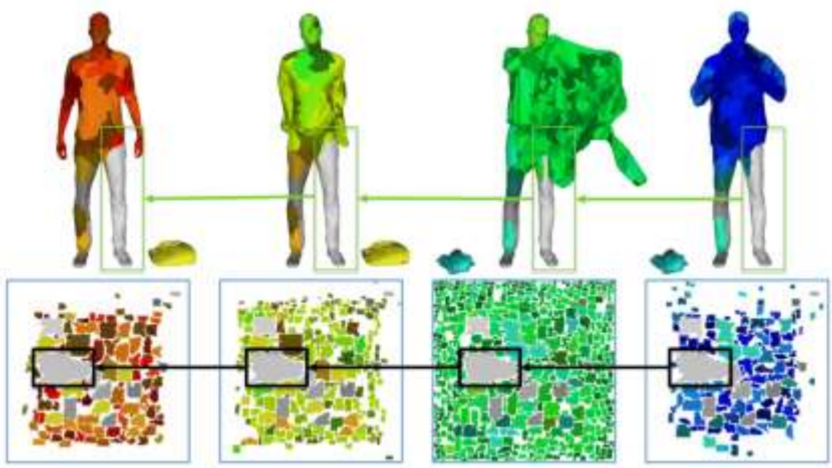

Figure 4: Mixing tracked and untracked content in the same model, as proposed in [8]

We believe that the concept can be extended even further, and mesh content could be mixed with point cloud content. Since most of the real-world acquired meshes come from reconstructing the mesh surface of a point cloud representation, parts of the object that might be better represented as point clouds could still remain point clouds, and the points that are well modeled by meshes could be converted. In this way, both data types could co-exist in the same model. Figure 5 shows a visualization of the concept, where the body is represented by a mesh, but the face is represented by point clouds, since it has finer details. One other possibility is also for zooming applications: for zoom out point cloud may be a better representation of certain desired regions and for zoomed in it could be better with mesh representation i.e. the content, depending on the application, may be morphed into different representations.

Figure 5: Mixed point cloud and mesh content.

54 Conclusion

We reported in this contribution several different mesh characteristics, (e.g. voxel/float geometry, registered/non-registered, noisy/clean) that we believe the group should consider moving forward with mesh compression standardization.

We suggest the group to discuss the use-cases and adopt the proposal of inclusion of different meshes in the requirements for mesh compression.

65 References

[1] https://www.sony.net/SonyInfo/technology/stories/Volumetric_Capture/

[2] https://www.quanticdream.com/en/beyond-two-souls

[3] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk and S. Sullivan, "High-quality streamable free-viewpoint video," in *ACM Transaction on Graphics (SIGGRAPH)*, 2015.

[4] https://www.cgtrader.com/free-3d-models?keywords=

[5] Dou, Mingsong, et al. "Fusion4d: Real-time performance capture of challenging scenes." ACM Transactions on Graphics (TOG) 35.4 (2016): 1-13.

[6] Newcombe, Richard A., Dieter Fox, and Steven M. Seitz. "Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[7] Guo, Kaiwen, et al. "The relightables: Volumetric performance capture of humans with realistic relighting." ACM Transactions on Graphics (TOG) 38.6 (2019): 1-19.

[8] Fabian Prada, Misha Kazhdan, Ming Chuang, Alvaro Collet, ´ and Hugues Hoppe. Spatiotemporal atlas parameterization for evolving meshes. ACM Trans. on Graphics (TOG), 2017.

[9] Soriano, Marc. "Skeletal Animation". Bourns College of Engineering. Retrieved 5 January 2011. http://alumni.cs.ucr.edu/~sorianom/cs134_09win/lab5.htm