**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2019/m49414**

**July 2019, Gothenburg, SE**

|  |  |
| --- | --- |
| **Source** | **Apple** |
| **Status** | **Input document** |
| **Title** | **Volumetric Tiling Information SEI message for V-PCC** |
| **Author** | **Alexis Michael Tourapis, Jungsun Kim, Khaled Mammou, David Singer, Fabrice Robinet** |

# Introduction

Spatial/"region of interest" random access and scalability are highly desirable for some point cloud applications. In order to achieve such functionality additional information may have to be provided within the bitstream that can be utilized by a decoder to determine which parts of the bitstream should be decoded and utilized to achieve the desired functionality. In particular, for region of interest decoding, the information signaled may include the area in the patch frame/atlas that contains all the relevant information to be decoded that corresponds to the region of interest, as well as the correspondence of this area in 3D space.

Given the temporal nature of a point cloud sequence, such information also needs to be very compact and it is highly desirable if the signaling mechanism for such information is simple enough and does not introduce too much overhead.

In this contribution we introduce a new SEI message, named as the volumetric tiling information SEI message, that could enable such a signaling mechanism. The method presented is inspired by an approach also used in the HEVC specification for the signaling of the Annotated Region SEI messages [2].

# Volumetric Tiling Information SEI message

To achieve the desired functionality, i.e. support region of interest/spatial random access or scalability within a V-PCC application, it is necessary to first appropriately partition the patch frame/atlas of a V-PCC frame into several segments, where each segment only contains patches that correspond to only certain regions of the reconstructed point cloud. An example is shown in Figure 1. In this example, 5 different tiling segments are indicated, each containing a different number of patches. These tiling segments, for simplicity are defined to be rectangular and need to be associated with also a bounding box in the 3D space. Labeling these segments might also be highly desirable, while it is also quite possible that some segments may overlap, e.g. because a particular segment in the point cloud may be associated with multiple regions of a point cloud.

The above requirements imply that we may need to maintain a list of tiling segments, and for each segment we would need to be able to signal its corresponding coordinates (bounding box) on the V-PCC atlas, but also its correspondence in the 3D space. An optional labeling. of the tiling segments may also be useful. However, efficient signaling of such information while considering the temporal characteristics of a V-PCC signal, is also highly desirable. For example, it is possible that an object remains within its 2D and/or its 3D bounding box for the entire sequence, or after a certain point in time that all or some of its 2D or 3D characteristics change. An object may also disappear and would need to be removed from the maintained tiling segment list.

Although we could have managed this process by defining a similar method as used for the inter predicted patches in V-PCC, i.e. by maintaining a tiling segment frame buffer and allow inter prediction using this buffer, that may be too complicated to manage especially for some applications. Instead, similar to the process defined for annotated regions, we define a process based on persistence of the information according to their output order and selective updating of information.



Figure . Atlas Tiling Segments containing patches that correspond to different 3D regions

In particular, we permit the signaling of up to N, e.g. 256, objects or tiling segments and we can assign to each one of these objects a set of 2D and 3D coordinates that define their relationship to the 2D atlas and the 3D point cloud representation space respectively. Optionally a label can also be assigned to the object. Then, this information and object can persist for all frames tha precede it in output order until a cancel flag is sent, which permits us to update some parameters of the object or even delete the object completely. New objects could also be introduced in time. The specifics of the syntax that we propose are shown in Section 3. The syntax in that section could also be extended with additional information, such as whether the object has temporarily disappeared from the scene instead of needing to be removed outright[[1]](#footnote-1), additional flags/parameters to signal partial updating of some information, e.g. only of the 2D or 3D size that corresponds to the object, association with patch tile groups, number and type of materials,the number of patches expected to be found in the area corresponding to the tiling segment, overall number of expected decoded points, intent of the tiling segment (e.g. random access versus scalability), and layer information for scalability purposes amongst others. We will let the topic of what additional parameters we should maybe consider to include in this SEI message as part of the decision process for its adoption.

# Volumetric Tiling Information SEI message syntax

|  |  |
| --- | --- |
| volumetric\_tiling\_info( payloadSize ) { | **Descriptor** |
| **vti\_cancel\_flag** | u(1) |
| if (!vti\_cancel\_flag) { |  |
| **vti\_object\_label\_present\_flag** | u(1) |
| **vti\_3d\_bounding\_box\_present\_flag** | u(1) |
| **vti\_object\_priority\_present\_flag** | u(1) |
| **vti\_object\_hidden\_present\_flag** | u(1) |
| **vti\_object\_collision\_shape\_present\_flag** | u(1) |
| **vti\_object\_dependency\_present\_flag** | u(1) |
| if( vti\_object\_label\_present\_flag ) { |  |
| **vti\_object\_label\_language\_present\_flag** | u(1) |
| if( vti\_object\_label\_language\_present\_flag ) { |  |
| while( !byte\_aligned( ) ) |  |
| **vti\_bit\_equal\_to\_zero** /\* equal to 0 \*/ | f(1) |
| **vti\_object\_label\_language** | st(v) |
| } |  |
| **vti\_num\_object\_label\_updates** | ue(v) |
| for( i = 0; i < vti\_num\_object\_label\_updates; i++ ) { |  |
| **vti\_label\_idx**[ i ] | ue(v) |
| **vti\_label\_cancel\_flag** | u(1) |
| LabelAssigned[ vti\_label\_idx[ i] ] = !vti\_label\_cancel\_flag |  |
| if ( !vti\_label\_cancel\_flag) { |  |
| while( !byte\_aligned( ) ) |  |
| **vti\_bit\_equal\_to\_zero** /\* equal to 0 \*/ | f(1) |
| **vti\_label**[ vti\_label\_idx[ i ] ] | st(v) |
| } |  |
| } |  |
| } |  |
| **vti\_num\_object\_updates** | ue(v) |
| if (vti\_3d\_bounding\_box\_present\_flag) { |  |
| **vti\_bounding\_box\_scale\_log2** | u(5) |
| **vti\_3d\_bounding\_box\_scale\_log2** | u(5) |
| **vti\_3d\_bounding\_box\_precision\_minus8** | u(5) |
| } |  |
| for( i = 0; i  <=  vti\_num\_object\_updates; i++ ) { |  |
| **vti\_object\_idx**[ i ] | ue(v) |
| **vti\_object\_cancel\_flag** | u(1) |
| ObjectTracked[**vti\_object\_idx**[ i ] ] = ! vti\_object\_cancel\_flag |  |
| if (!vti\_object\_cancel\_flag) { |  |
| **vti\_bounding\_box\_update\_flag** | u(1) |
| if( vti\_bounding\_box\_update\_flag ) { |  |
| **vti\_bounding\_box\_top[** vti\_object\_idx[ i ] ] | u(v) |
| **vti\_bounding\_box\_left**[ vti\_object\_idx[ i ] ] | u(v) |
| **vti\_bounding\_box\_width**[ vti\_object\_idx[ i ] ] | u(v) |
| **vti\_bounding\_box\_height**[ vti\_object\_idx[ i ] ] | u(v) |
| **}** |  |
| if( vti\_3d\_bounding\_box\_present\_flag ) { |  |
| **vti\_3d\_bounding\_box\_update\_flag** | u(1) |
| if( vti\_3d\_bounding\_box\_update\_flag) { |  |
| **vti\_3d\_bounding\_box\_x[** vti\_object\_idx[ i ] **]** | u(v) |
| **vti\_3d\_bounding\_box\_y[** vti\_object\_idx[ i ] **]** | u(v) |
| **vti\_3d\_bounding\_box\_z[** vti\_object\_idx[ i ] **]** | u(v) |
| **vti\_3d\_bounding\_box\_delta\_x[** vti\_object\_idx[ i ] **]** | u(v) |
| **vti\_3d\_bounding\_box\_delta\_y[** vti\_object\_idx[ i ] **]** | u(v) |
| **vti\_3d\_bounding\_box\_delta\_z[** vti\_object\_idx[ i ] **]** | u(v) |
| } |  |
| } |  |
| if (vti\_object\_priority\_present\_flag) { |  |
| **vti\_object\_priority\_update\_flag** | u(1) |
| if (vti\_object\_priority\_update\_flag) { |  |
| **vti\_object\_priority\_value[** vti\_object\_idx[ i ] **]** | u(4) |
| **}** |  |
| **}** |  |
| if (vti\_object\_hidden\_present\_flag) { |  |
| **vti\_object\_hidden\_flag**[ vti\_object\_idx[ i ] **]** | u(1) |
| **}** |  |
| if( vti\_object\_label\_present\_flag ) { |  |
| **vti\_object\_label\_update\_flag** | u(1) |
| if( vti\_object\_label\_update\_flag ) |  |
| **vti\_object\_label\_idx**[ vti\_object\_idx[ i ] **]** | ue(v) |
| } |  |
| if (vti\_object\_collision\_shape\_present\_flag) { |  |
| **vti\_object\_collision\_shape\_update\_flag** | u(1) |
| if (vti\_object\_collision\_shape\_update\_flag) { |  |
| **vti\_object\_collision\_shape\_id**[ vti\_object\_idx[ i ] ] | u(16) |
| **}** |  |
| **}** |  |
| if (vti\_object\_dependency\_present\_flag) { |  |
| **vti\_object\_dependency\_update\_flag** | u(1) |
| if (**vti\_object\_dependency\_update\_flag**) { |  |
| **vti\_object\_num\_dependencies**[ vti\_object\_idx[ i ] ] | u(4) |
| for( j = 0; j  <  vti\_object\_num\_dependencies[ vti\_object\_idx[ i ] ]; j++ ) { |  |
| **vti\_object\_dependency\_idx**[ vti\_object\_idx[ i ] ] [ j ] | u(8) |
| **}** |  |
| **}** |  |
| **}** |  |
| } |  |
| } |  |
| } |  |
| } |  |

# Conclusion

A new SEI message, the Volumetric Tiling Information SEI message, was proposed in this document. This SEI message provides a flexible mechanism for indicating point cloud tiling segments in the context of V-PCC and their correspondence in 3D space. It is suggested that this SEI message is studied further and potentially adopted by the 3DG group in the V-PCC specification.

# References

1. “Study Text of ISO/IEC CD 23090-5 Video-based Point Cloud Compression”, ISO/IEC JTC1/SC29 WG11 (MPEG) output document N18180, Geneva, May 2019.
2. "Annotated regions and fisheye video information SEI messages for HEVC", JCTVC-AH1012, 34th JCT-VC meeting, Marrakesh, MA, Jan 2019.

1. This could also be done by setting one or all of the bounding box size information to 0. [↑](#footnote-ref-1)