ISO/IEC JTC 1/SC 29/WG 03 N1325

**ISO/IEC JTC 1/SC 29/WG 03  
MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title:** Technologies under consideration on carriage of V3C data

**Status:** Approved

**Date of document:** 2024-08-27

**Source:** ISO/IEC JTC 1/SC 29/WG 03

**No. of pages:** 119 (with cover page)

**Email of Convenor:** young.L @ samsung . com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 03 MPEG SYSTEMS**

**ISO/IEC JTC 1/SC 29/WG 03 N1325**

**July 2024, Sapporo, JP**

|  |  |
| --- | --- |
| **Title** | **Technologies under consideration on carriage of V3C data** |
| **Source** | **WG 03, MPEG Systems** |
| **Status** | **Approved** |
| **Serial Number** | **4245** |

# Introduction

This document contains technologies under consideration for carriage of V3C data as of the 16th MPEG Systems WG meeting. As 2nd edition of 23090-10 is being developed, the technologies in this document will be migrated to the WD of the 2nd edition at the 17th meeting unless no major open issues are identified.

# Support of 2D snapshot (m66538)

## Introduction

Compressing volumetric contents has strong benefit for saving resources for storage and delivery of the contents. However, it introduces a challenge for quick preview or trick play of the contents similar to any compressed video data has. Several video codecs should be initiated and more than one video frames for each components of a compressed volumetric frame should be decoded and a volumetric frame should be reconstructed with the decoded results for a quick preview of volumetric content. As one directional or bidirectional dependent coding could have been also applied to further enhance compression efficiency, more than one video frame should be decoded to get a specific frame of volumetric content. If random access points of the components are not aligned or the frame rates of the components are different each other, a greater number of video frames should be decoded to get the result. So, quick preview or trick play of a volumetric contents which used be a straight forward easy job for uncompressed volumetric contents become quite complicated resource and time-consuming thing when the contents are compressed.

To solve such issues and make quick preview or trick play operation simpler and easier 2D snapshot images of volumetric contents at certain points of time with a camera at a certain position and direction could be provided. A client can decode and present 2D snapshot images instead of volumetric contents by decoding compressed bitstreams and compositing/rendering volumetric contents when it is not really needed.

This contribution proposes a method to add one or more tracks to carry 2D snapshot images for a volumetric content.

## 2D snapshot image track

### Overview

The 2D snapshot image track contains one or more samples of coded bitstream of 2D image of a coded volumetric frame rendered at a certain location and direction. Each sample contains a 2D projected image of a coded volumetric frame whose composition time is same with such sample. There can be more than one 2D snapshot image track for a single CVS and each of them contains different version of snapshot image.

### Restriction to the track

The value of handler\_type of the 2D snapshot image track shall be ‘vide.’ In other words, the track used as a 2D snapshot image should be the one which could have ‘vide’ as a value for handler\_type. All samples in the 2D snapshot image track shall be sync samples.

### Track references

To associate a 2D snapshot image track with the tracks containing V3C data, track reference tool of ISO/IEC 14496-12 shall be used. One or more TrackReferenceTypeBoxes shall be added to a TrackReferenceBox within the TrackBox of the V3C atlas track or V3C atlas tile track, one for each 2D snapshot image tracks. The TrackReferenceTypeBox shall contain array of track\_IDs designating the tracks containing 2D snapshot images which the V3C atlas track or V3C atlas tile track references. The 4CC value of reference\_type of such TrackReferenceTypeBox shall be ‘2dsi.’

### Indication of camera used for rendering snapshot images

Information about the camera used to render 2D snapshot images is provided as viewport information timed-metadata track. A viewport sample whose composition time is same with a snapshot image provide information about the camera used to render such image. When such viewport timed-metadata track is provided the value of viewport\_type is set to ‘0.’ The viewport information timed-metadata track shall reference corresponding 2D snapshot image track instead of V3C atlas track and ‘2dci’ shall be used for reference\_type.

## Conclusion

It is proposed to add a method to carry the 2D snapshot images to support simple and easy preview or scanning of compressed volumetric contents. By adding this feature compressed volumetric contents will also become useful for the use cases where decoding, composition and rendering of compressed volumetric contents is challenging.

# Support of multi-directional 2D snapshot (m67720)

## Introduction

Compressing volumetric contents has strong benefit for saving resources for storage and delivery of the contents. However, it introduces a challenge for quick preview or trick play of the contents similar to any compressed video data has. Several video codecs should be initiated and more than one video frames for each components of a compressed volumetric frame should be decoded and a volumetric frame should be reconstructed with the decoded results for a quick preview of volumetric content. As one directional or bidirectional dependent coding could have been also applied to further enhance compression efficiency, more than one video frame should be decoded to get a specific frame of volumetric content. If random access points of the components are not aligned or the frame rates of the components are different each other, a greater number of video frames should be decoded to get the result. So, quick preview or trick play of a volumetric contents which used be a straight forward easy job for uncompressed volumetric contents become quite complicated resource and time-consuming thing when the contents are compressed.

To solve such issues and make quick preview or trick play operation simpler and easier 2D snapshot images of volumetric contents at certain points of time with a camera at a certain position and direction could be provided. A client can decode and present 2D snapshot images instead of volumetric contents by decoding compressed bitstreams and compositing/rendering volumetric contents when it is not really needed.

At the previous meeting, a method to add one or more tracks to carry 2D snapshot images for a volumetric content has been presented. The contribution proposes to provide a track referenced from a volumetric content track containing 2D snapshot images. It proposes to optionally include viewport information timed-metadata track to indicate information about the camera use to render the 2D snapshot images as [m66538](https://dms.mpeg.expert/doc_end_user/documents/145_OnLine/wg11/m66538-v2-m66538v22Dsnapshotimagetrack.zip).

In this contribution, a method to provide more than one 2D snapshot image for a certain coded volumetric frame is proposed. As a volumetric frame can be viewed from multiple directions a single snapshot image may not be sufficient to provide enough information about a volumetric frame. It would be possible to provide more than one snapshots by adding more than one snapshot image tracks but it would be not quite efficient for certain use case, e.g. camera is fixed for entire volumetric frames in a track. As modern video codecs such as HEVC, VVC or EVC provides mechanisms to include more than one subset of pictures such as tiles or subpictures, and ISO/IEC 14496-15 can indicate each of them as a separate sub-sample, this contribution proposes a method to include more than one snapshot images by using such features when the camera information is same for entire samples in a track.

## Multi-directional snapshot image track

### Overview

The multi-directional snapshot image track contains one or more samples of coded bitstream of 2D image of a coded volumetric frame. Each sample consist of more than one sub-samples and each sub-sample contains a 2D projected image of a coded volumetric frame where location and direction of projection of each sub-sample are not same each other. SubSampleInformationBox shall present in a multi-directional snapshot image track to provide subsample information. The set of location and direction of cameras used for projection shall remain same for a single track. Composition time of the 2D images shall be same with the sample of coded volumetric frame.

### Track references

To associate a multi-directional snapshot image track with the tracks containing V3C data, track reference tool of ISO/IEC 14496-12 shall be used. One or more TrackReferenceTypeBoxes shall be added to a TrackReferenceBox within the TrackBox of the V3C atlas track or V3C atlas tile track, one for each 2D snapshot image tracks. The TrackReferenceTypeBox shall contain array of track\_IDs designating the tracks containing 2D snapshot images which the V3C atlas track or V3C atlas tile track references. The 4CC value of reference\_type of such TrackReferenceTypeBox shall be ‘mdsi.’

### Restriction to the track

The track referenced from a V3C atlas track or V3C atlas tile track with reference\_type ‘mdsi shall be represented in the file as restricted video and shall use a generic sample entry ‘resv’ with following additional requirements:

— SchemeTypeBox shall be present in RestrictedSchemeInfoBox and scheme\_type is set to 'mdst'

— All samples in the track shall be sync samples

— SubSampleInformationBox shall be present and the value of subsample\_count of SubSampleInformationBox shall be same for all entries and greater than one.

#### Multi-dimensional snapshot camera information box

##### Definition

Box Type: 'mdst'

Container: SchemeInformationBox

Mandatory: Yes (when the SchemeType is 'mdst')

Quantity: One

The Multi-dimensional snapshot camera information box is used to indicate the information about the camera used to render the snapshots of volumetric frames for each sub-samples indicated by SubSampleInformationBox. i-th view port information shall indicate i-th sub-sample in bitstream order within a sample.

##### Syntax

aligned(8) class MultiDimSnapshotCameraInfoBox extends extends FullBox('mdst', version = 0, 0)

{

unsigned int(8) num\_viewports;

for (int i=1; i <= num\_viewports; i++){

unsigned int(1) camera\_extrinsic\_flag[i];

unsigned int(1) camera\_intrinsic\_flag[i];

bit(6) reserved = 0;

ViewportInfo (camera\_extrinsic\_flag[i], camera\_intrinsic\_flag[i]);

}

}

##### Semantics

num\_viewport indicates the number of viewport signaled in the sample. The value of this field shall be equal to the value of the value of subsample\_count of SubSampleInformationBox. i-th viewport provides information about the camera for the i-th subsample in bitstream order.

camera\_intrinsic\_flag[i] equal to 1 indicates that the intrinsic camera parameters are present in the i-th viewport.

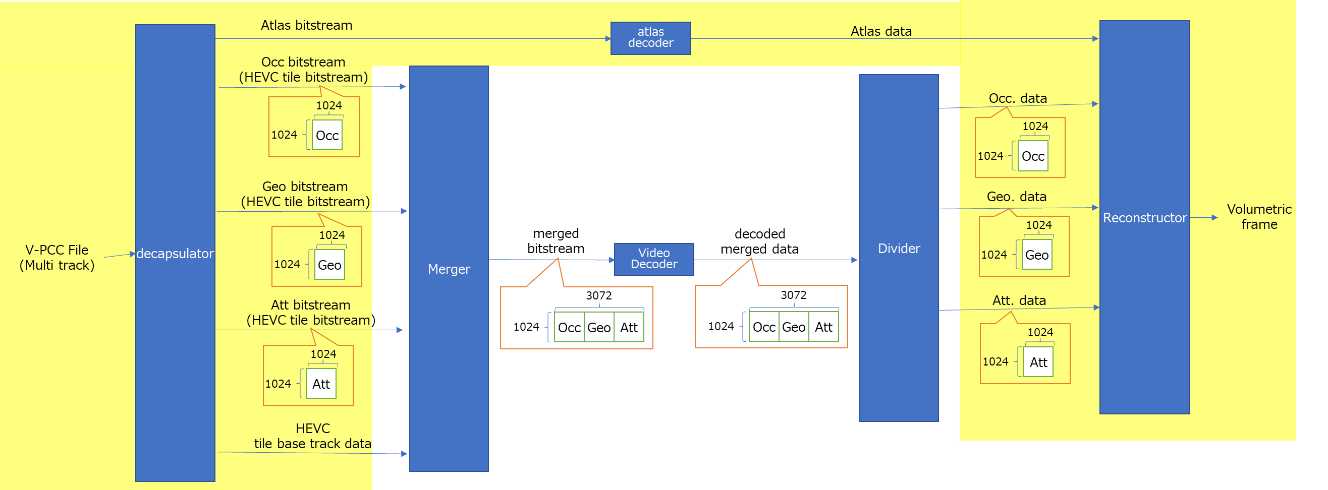
camera\_extrinsic\_flag[i] equal to 1 indicates that the extrinsic camera parameters are present in the i-th viewport.

ViewportInfo provides information about the camera. Then syntax and semantics of this class is specified in subclause 10.2.3 of ISO/IEC 23090-10

# Supporting to decode with single decoder instance for V-PCC content (m67634)

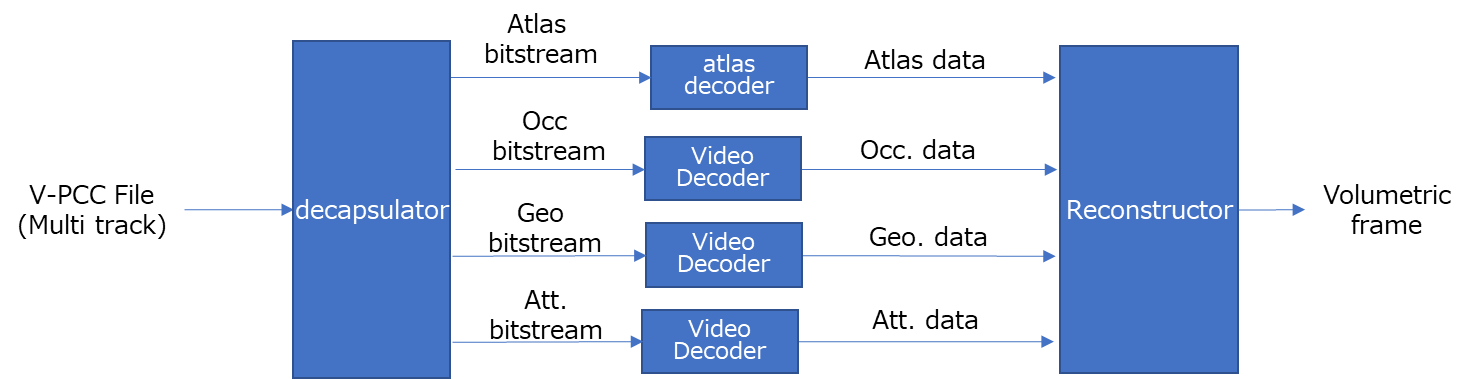
## Proposal

In this contribution, it proposes following:

1. Update the process diagram which include a non-video component, for comment resolution.  
   Add an Atlas bitstream and a reconstruction process using decoded atlas data and each decoded video data to each implementation example diagram.  
   Figure 1 shows an example of a complete decoder structure using HEVC tiles. The added part is highlighted in yellow. It also updates figures which using VDI operation and using VVC subpicture in the same way.  
     
   

**Figure 1 complete decoder structure using HEVC tiles**

1. Adds the basic decoding V-PCC content process with multiple video decoder instance because it helps the readers understand by clarifying the difference from decoding process with single video decoder instance. Specifically, it adds the Figure 2 and process description in the ‘General’ clause.



**Figure 2 complete decoder structure for V-PCC content**

## Proposal text

The updates as explained in the section 2 are integrated into previous proposal text [2] and are highlighted in yellow.

-------

Add the Bibliography.

[3] ISO/IEC 23090-13:202x , Information technology — Coded representation of immersive media — Part 13: Video decoding interface for immersive media

Add the new Annex.

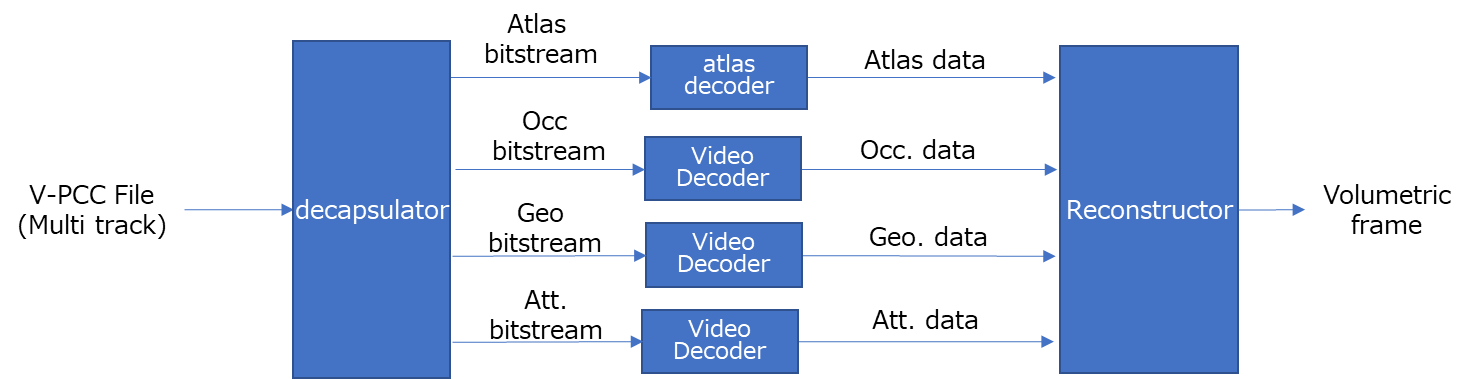
**Annex X (informative) Implementation examples of decoding all video components of V3C contents with single decoder instance**

**X.1 General**

For V3C content with multiple video components, each video component will be decoded individually.

For example, Figure X.1 shows the processing when V-PCC file which multi-track encapsulated atlas, occupancy, geometry, and attribute encoded data.

V-PCC file is decapsulated to atlas, occupancy, geometry, and attribute bitstream in decapsulator, respectively All bitstream are decoded in decoder, respectively. It is reconstructed into a volumetric frame using all decoded data. This case is typically handled by multiple video decoder instances.



**Figure X.1 decoding process of V-PCC content**

This annex describes implementation examples for decoding all video components of V3C content with single video decoder instance using function defined in other specifications.

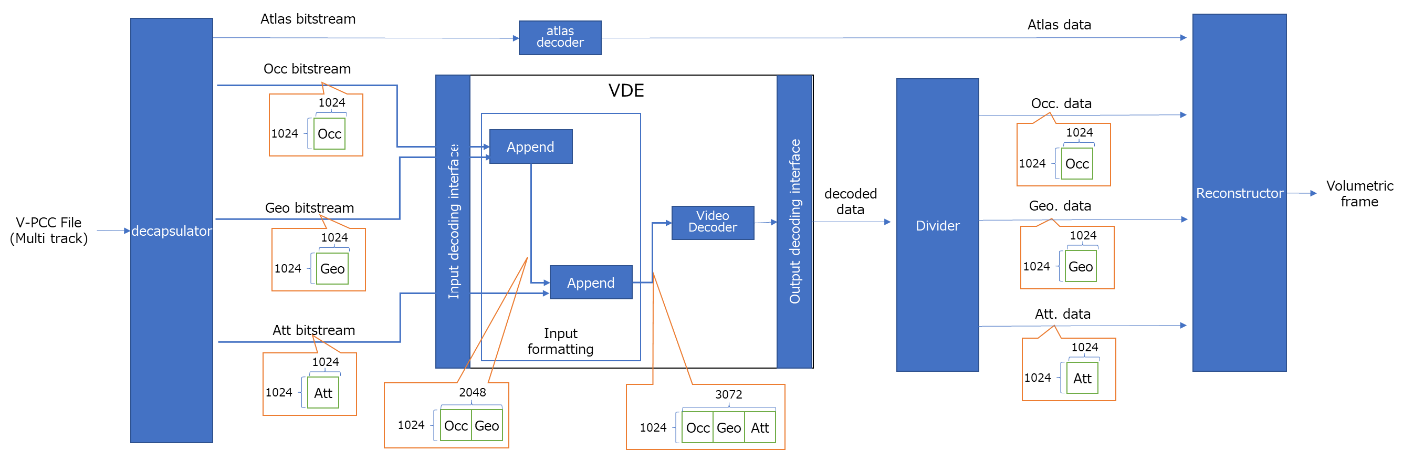
**X.2 Example of decoding with a single decoder instance using ISO/IEC 23090-13**

ISO/IEC 23090-13 Video decoding interface specifies the interface of Video Decoding Engine (VDE) and operation of elementary stream in VDE.

The “inserting” function is one of the specified operations in ISO/IEC 23090-13 and it can insert an elementary stream into another elementary stream and generate one merged elementary stream.

An implementation example is shown in Figure X.2. A V-PCC file which multi-track encapsulated atlas, occupancy, geometry, and attribute encoded data is decapsulated to atlas, occupancy, geometry, and attribute bitstream in decapsulator, respectively. The occupancy, geometry, and attribute bitstream are input to the VDE. The first “inserting” function of VDI operation insert the geometry bitstream into the occupancy bitstream. Then, the second “inserting” function insert the attribute bitstream into the bitstream which generated by the first “inserting” function. As a result, generated bitstream (as merged bitstream in the figure) contains occupancy, geometry, and attribute. The bitstream contained occupancy, geometry, and attribute is decoded with single decoder instance. The decoded data which is output from decoder is divided into occupancy, geometry, and attribute data. These data and atlas data are reconstructed to the volumetric frame.

Note: the “inserting” function needs to regenerate parameter sets to ensure the bitstream conformance.

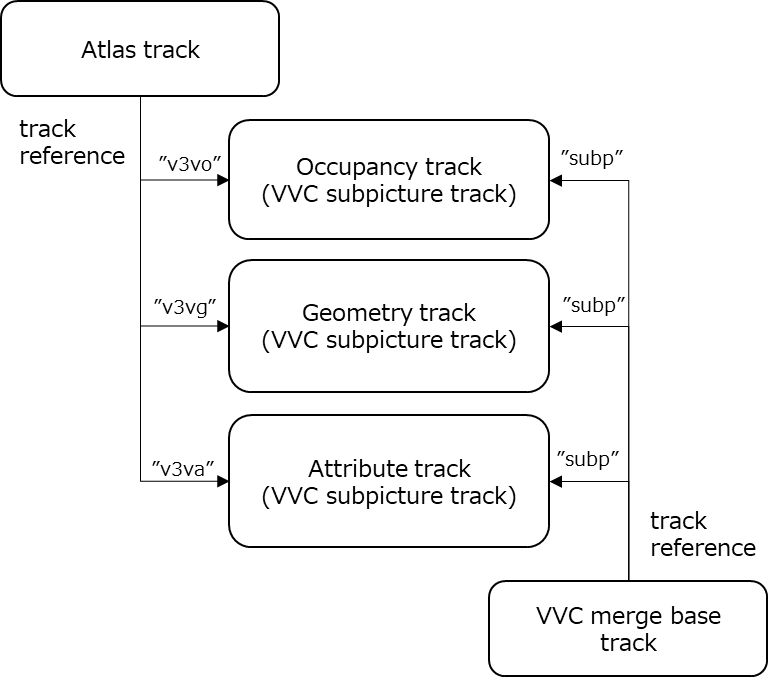
****

**Figure X.2 decoding process using ISO/IEC 23090-13**

**X.3 Example of decoding with a single decoder instance using bitstream reconstruction of VVC subpicture**

It specifies the bitstream reconstruction using independent VVC subpictures and VVC merge base tracks in the ISO/IEC 14496-15.

In this example, the V3C video component is encoded as an independent VVC subpicture. For example, all V3C video components may be encoded as independent subpictures in one picture, and then divide into tracks for each V3C video component(VVC subpicture). And the VVC merge base track is generated to merge VVC subpicture tracks which are occupancy, geometry, attribute video component track. Figure X.3 shows the structure of VVC content. The “subp” track reference in VVC merge base track refers to the occupancy, geometry, attribute video component track.

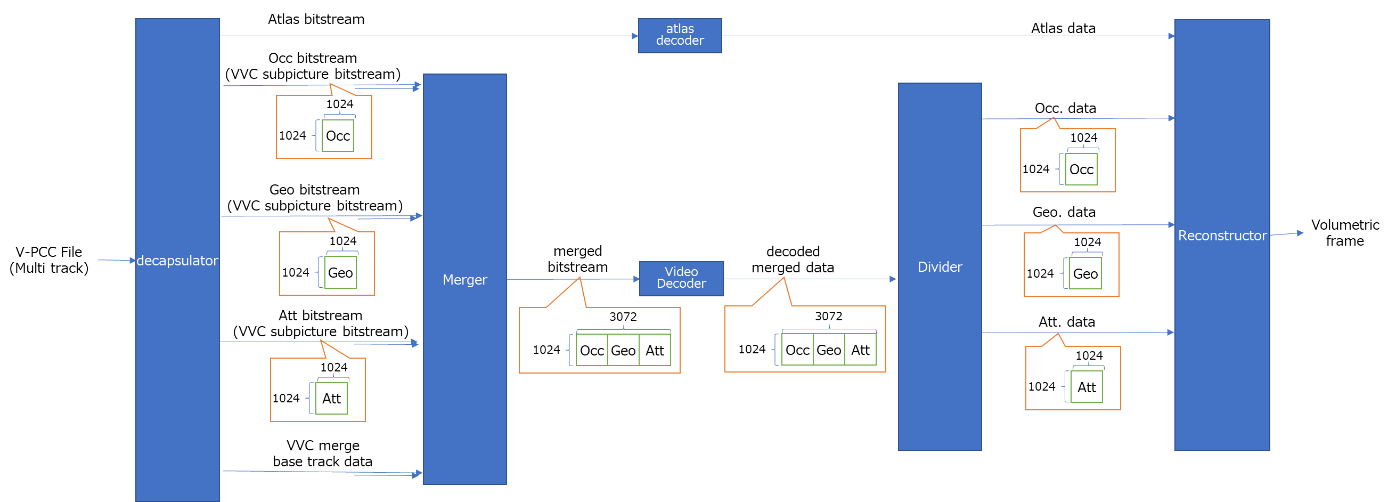


**Figure X.3 track structure of V3C content using bitstream reconstruction of VVC subpicture**

When the client plays the V3C content, it may identify the VVC merge base track after finding the atlas track. To identify the VVC merge base track, the client may choos the VVC base track that the tracks referenced by the ‘subp’ track reference in the VVC base track are identical the video component tracks referenced from the track reference in the atlas track.

The decoding process of V-PCC file which multi-track encapsulated atlas, occupancy, geometry, and attribute encoded data is shown in Figure X.4. V-PCC file which multi-track encapsulated atlas, occupancy, geometry, and attribute encoded data is decapsulated to atlas, occupancy, geometry, and attribute bitstream in decapsulator, respectively.. The merged bitstream is generated from VVC merge base track and VVC subpicture tracks which are occupancy, geometry, attribute video component track according to bitstream reconstruction defined in ISO/IEC 14496-15.

The bitstream (as merged bitstream in Figure X.4) contained occupancy, geometry, and attribute is decoded with a single decoder instance. The decoded merged data which is output from decoder is divided into occupancy, geometry, and attribute data using the information of placement and resolution in ‘trif’ sample group in the VVC merge base track. These divided data and atlas data are reconstructed to the volumetric frame.



**Figure X.3 decoding process using bitstream reconstruction of VVC subpicture**

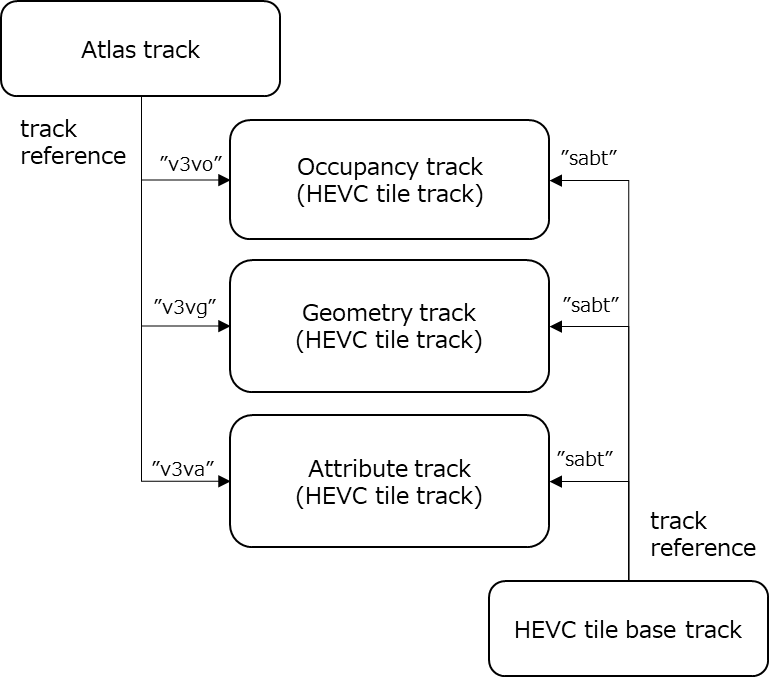
Note1 : The VVC picture is rectangular picture. In other words, the merged picture is rectangular picture. To conformant this condition, it need to consider the resolution of each video component and placement of each video component in the merged picture. In the Figure X.3, all video components are the same resolution, And the merged picture can be rectangular picture by arranging from pictures horizontally.

Note2 : The placement of subpictures in the merged bitstream is configured in the VVC merge base track. The placement of subpictures may be arranged not only horizontally, but also vertically, or both horizontally and vertically

**X.4 Example of decoding with a single decoder instance using bitstream reconstruction of HEVC tile**

It specifies the bitstream reconstruction using independent HEVC tiles and HEVC tile base tracks in the ISO/IEC 14496-15.

In this example, the V3C video component is encoded as an independent HEVC tile. For example, all V3C video components may be encoded as independent HEVC tile in one picture, and then divide into tracks for each V3C video component (HEVC tile). And the HEVC tile base track is generated to merge HEVC tile tracks which are occupancy, geometry, attribute video component track. Figure X.5 shows the structure of V3C content. The “sabt” track reference in VVC merge base track refers to the occupancy, geometry, attribute video component track.

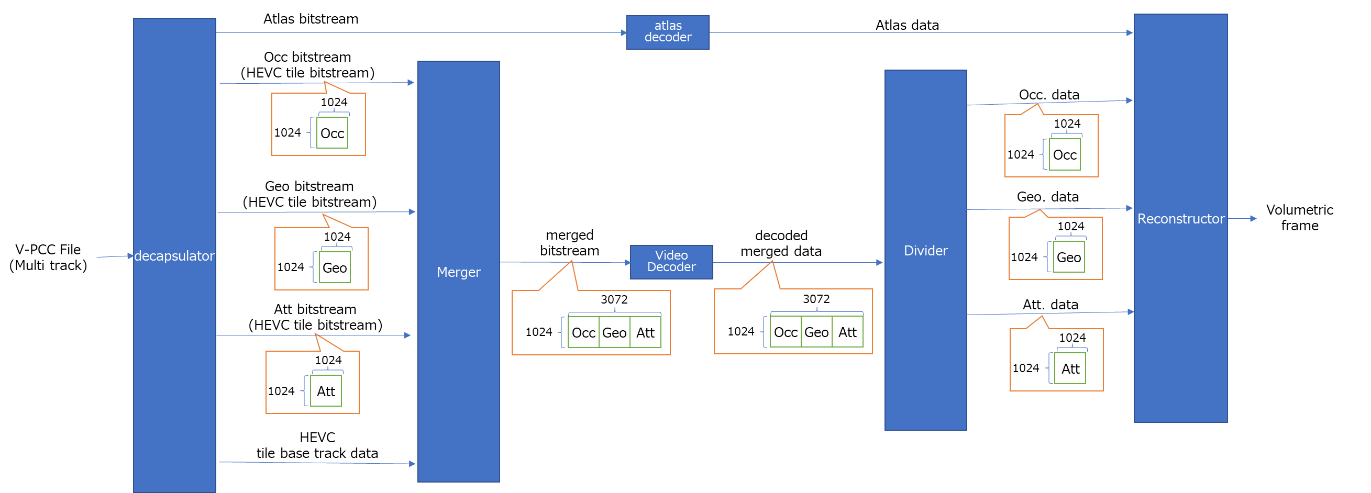


**Figure X.5 track structure of V3C content using bitstream reconstruction of HEVC tile**

When the client plays the V3C content, it can identify the HEVC tile base track after finding the atlas track. To identify the HEVC tile base track, the client can choos the HEVC tile track that the tracks referenced by the ‘sabt’ track reference in the VVC base track are identical the video component tracks referenced from the track reference in the atlas track.

The decoding process of V-PCC file which multi-track encapsulated atlas, occupancy, geometry, and attribute encoded data is shown in Figure X.6. V-PCC file which multi-track encapsulated atlas, occupancy, geometry, and attribute encoded data is decapsulated to atlas, occupancy, geometry, and attribute bitstream in decapsulator, respectively. The merged bitstream is generated from HEVC tile base track and HEVC tile tracks which are occupancy, geometry, attribute video component track according to bitstream reconstruction defined in ISO/IEC 14496-15.

The bitstream (as merged bitstream in Figure X.6) contained occupancy, geometry, and attribute is decoded with a single decoder instance. The decoded sequence which is output from decoder is divided into occupancy, geometry, and attribute sequences using the information of placement and resolution in ‘trif’ sample group in the HEVC tile base track. These divided data and atlas data are reconstructed to the volumetric frame.



**Figure X.6 decoding process using bitstream reconstruction of HEVC tile**

Note1 : The HEVC picture is rectangular picture. In other words, the merged picture is rectangular picture. To conformant this condition, it need to consider the resolution of each video component and placement of each video component in the merged picture. For example, if all video components are the same resolution, the merged picture can be rectangular picture by arranging from pictures horizontally (or vertically).

Note2 : The placement of HEVC tile in the merged bitstream is configured in the HEVC base track. The placement of subpictures may be arranged not only horizontally, but also vertically, or both horizontally and vertically.

# References

1. m66538, “[38.2] 2D snapshot image track for V3C content,” MPEG145, January 2024
2. m66720, “[37.2] Supporting multi-directional 2D snapshot for V3C content,” MPEG146, April 2024
3. m67634, “[VOL-SYS] Supporting to decode with single decoder instance for V-PCC content,” MPEG146, April 2024