frontiers | Frontiers in Signal Processing

TECHNOLOGY REPORT
published: 29 April 2022
doi: 10.3389/frsip.2022.883943

An Overview of the MPEG Standard for Storage and Transport of Visual Volumetric Video-Based Coding

Lauri Ilola[1]*, *Lukasz Kondrad*[1], *Sebastian Schwarz*[1] and *Ahmed Hamza*[2]

[1]Nokia Solutions and Networks GmbH & Co. KG, Munich, Germany, [2]InterDigital Canada Ltée, Montréal, QC, Canada

The increasing popularity of virtual, augmented, and mixed reality (VR/AR/MR) applications is driving the media industry to explore the creation and delivery of new immersive experiences. One of the trends is volumetric video, which allows users to explore content unconstrained by the traditional two-dimensional window of director's view.

The ISO/IEC joint technical committee 1 subcommittee 29, better known as the Moving Pictures Experts Group (MPEG), has recently finalized a group of standards, under the umbrella of Visual Volumetric Video-based Coding (V3C). These standards aim to efficiently code, store, and transport immersive content with 6 degrees of freedom. The V3C family of standards currently consists of three documents: 1) ISO/IEC 23090-5 defines the generic concepts of volumetric video-based coding and its application to dynamic point cloud data; 2) ISO/IEC 23090-12 specifies another application that enables compression of volumetric video content captured by multiple cameras; and 3) ISO/IEC 23090-10 describes how to store and deliver V3C compressed volumetric video content. Each standard leverages the capabilities of traditional 2D video coding and delivery solutions, allowing for re-use of existing infrastructures which facilitates fast deployment of volumetric video.

This article provides an overview of the generic concepts of V3C, as defined in ISO/IEC 23090-5. Furthermore, it describes V3C carriage related functionalities specified in ISO/IEC 23090-10 and offers best practices for the community with respect to storage and delivery of volumetric video.

Keywords: MPEG, V3C, V-PCC, MIV, ISOBMFF, DASH, virtual reality

OPEN ACCESS

Edited by:
Matteo Naccari,
Audinate, United Kingdom

Reviewed by:
Jesús Gutiérrez,
Universidad Politécnica de Madrid, Spain
Maria Paula Queluz,
Universidade de Lisboa, Portugal

***Correspondence:**
Lauri Ilola
lauri.ilola@nokia.com

Specialty section:
This article was submitted to Image Processing, a section of the journal Frontiers in Signal Processing

Received: *25 February 2022*
Accepted: *11 April 2022*
Published: *29 April 2022*

Citation:
Ilola L, Kondrad L, Schwarz S and Hamza A (2022) An Overview of the MPEG Standard for Storage and Transport of Visual Volumetric Video-Based Coding. Front. Sig. Proc. 2:883943. doi: 10.3389/frsip.2022.883943

1 INTRODUCTION

Unlike a three degrees of freedom (3DoF) experience, an immersive six degrees of freedom (6DoF) representation enables a larger viewing-space, wherein viewers have both translational and rotational freedom of movement. In a 3DoF visual experience, content is presented to viewers as if they were positioned at the center of a scene, looking outwards, with all parts of the content positioned at a constant distance. 6DoF experiences allow viewers to move freely in the scene and experience the content from various viewpoints. Contrarily to 3DoF, 6DoF videos enable perception of motion parallax, where the change in relative geometry between objects is reflected with the pose of the viewer. The absence of motion parallax in 3DoF videos is inconsistent with the workings of a normal human visual system and often leads to visual discomfort (Kongsilp and Dailey, 2017).

The introduction of unconstrained viewer translation and motion parallax increases the amount of data required to describe the volumetric scene. Hence, the Motion Picture Experts Group (MPEG) has specified the Visual Volumetric Video-based Coding (V3C) standard ISO/IEC 23090-5 (ISO/IEC 23090-5, 2021) to efficiently code dynamic volumetric visual scenes. This standard caters to virtual reality, augmented reality, and mixed reality applications, such as gaming, sports broadcasting, motion picture productions, and telepresence (Schwarz et al., 2019).

The V3C standard defines a generic mechanism for coding volumetric video and can be used by applications targeting different flavors of volumetric content, such as point clouds, immersive video with depth, or even mesh representations of visual volumetric frames. So far, MPEG has specified two applications that utilize V3C: video-based point cloud compression (V-PCC), also specified in ISO/IEC 23090-5 (ISO/IEC 23090-5, 2021), and MPEG immersive video (MIV) specified in ISO/IEC 23090-12 (ISO/IEC 23090-12, 2021). For detailed information on the applications and how they apply V3C coding, the reader is referred to (Graziosi et al., 2020) and (Boyce et al., 2021) for V-PCC and MIV, respectively.

To enable storage and delivery of compressed volumetric content, MPEG developed a standard: carriage of visual volumetric video-based coding data ISO/IEC 23090-10 (ISO/IEC 23090-10, 2021). Like V3C coding standards, the systems aspects for volumetric content leverage existing technologies and frameworks for traditional 2D video. The ISO/IEC 23090-10 standard defines how V3C-coded content may be stored in an ISO base media file format (ISOBMFF) (ISO/IEC 14496-12, 2020) container as timed and non-timed data, providing the ability to multiplex V3C media with other types of media such as audio, video, or image. Moreover, the standard defines extensions to the Dynamic Adaptive Streaming over Hypertext Transfer Protocol (HTTP) (DASH) (ISO/IEC 23009-1, 2019) and MPEG Media Transport (MMT) (ISO/IEC 23008-1, 2017) frameworks to enable delivery of V3C-coded content over a network leveraging existing multimedia delivery infrastructures.

This article is organized as follows. Background information on V3C, ISOBMFF, and DASH is provided in **Section 2**. In **Section 3**, new elements introduced to ISOBMFF that enable storage of V3C content are described. Information about V3C streaming with focus on the utilization of DASH is provided in **Section 4**. In **Section 5**, we draw a conclusion and look at future standardization projects related to volumetric video in MPEG and other international standardization bodies.

2 BACKGROUND

2.1 Visual Volumetric Video-Based Coding (V3C)

Traditionally, video is seen as a series of still images which, when composed together, display a moving picture. Each still image or frame of the series is a two-dimensional grid of pixels. Similarly, volumetric video can be described as series of still volumetric frames, where each volumetric frame is a three-dimensional grid of voxels. The three-dimensional voxel grid can be represented in multiple forms. One example is a point cloud representation, where each point is associated with a geometry position together with the associated attribute information such as color, reflectance, transparency, etc.

Coding of 2D video has been studied for decades. As an example the first version of the H.261 video codec was specified in 1988 (Recommendation H.261, 1988). Many successful video codecs have since been implemented in billions of devices that are used every day. Coding of volumetric video, however, is still in its infancy. Visual Volumetric Video-based Coding (V3C) compresses volumetric video by taking advantage of the performance and ubiquity of traditional 2D video coding technologies. To achieve this, each volumetric frame is transformed from its 3D representation into multiple 2D representations and associated metadata known as *atlas data* in the V3C specification. After the conversion from 3D to 2D, the resulting 2D representations are compressed using traditional video codecs while atlas data are compressed with a separate encoding mechanism defined in ISO/IEC 23090-5 (ISO/IEC 23090-5, 2021).

Figure 1 illustrates at a high-level the process of converting the 3D representation into a 2D representation in a V3C encoder. The conversion is achieved by projecting the 3D representation of a volumetric frame (A) to a number of 2D planes, called patches (C), through a virtual camera (B). The patches are then arranged into 2D collections of patches creating the 2D frame representation (D). Each patch can have representations of geometry, attribute, and occupancy information. The geometry patches contain depth values indicating the distances of the projected points from the virtual camera plane, whereas the attribute patches represent the characteristics of the points, e.g. the color. The occupancy patches specify the validity of the points for reconstruction purposes. The information about the position of the patches, their dimensions, and the order in which they were packed in 2D frame representations is stored in the atlas data. Additionally, atlas data contain information about the virtual cameras and their intrinsic and extrinsic parameters used for generating the patches. Using the atlas data, a V3C decoder can perform the inverse projection from the 2D frame representations back to the 3D representation of the volumetric video, a process known as 3D reconstruction. During the reconstruction each valid pixel in a geometry patch, as indicated by the occupancy information, is re-projected back into 3D space as a point based on the depth value and the atlas data. The reconstructed point can be then textured based on the corresponding pixel in the attribute patch.

The maximum dimensions of the 2D frame representation of a V3C content depend on the used video codec; as commercially deployed decoders are typically constrained in terms of video resolution and frame rate. To circumvent these limitations, V3C allows splitting the projected patches into multiple 2D frame representations and corresponding associated metadata, thus creating multiple atlases. To avoid duplicating data, such as projection parameters, between multiple atlases, V3C defines a common atlas data structure, which contains information that applies for all atlases of the presentation. For more information

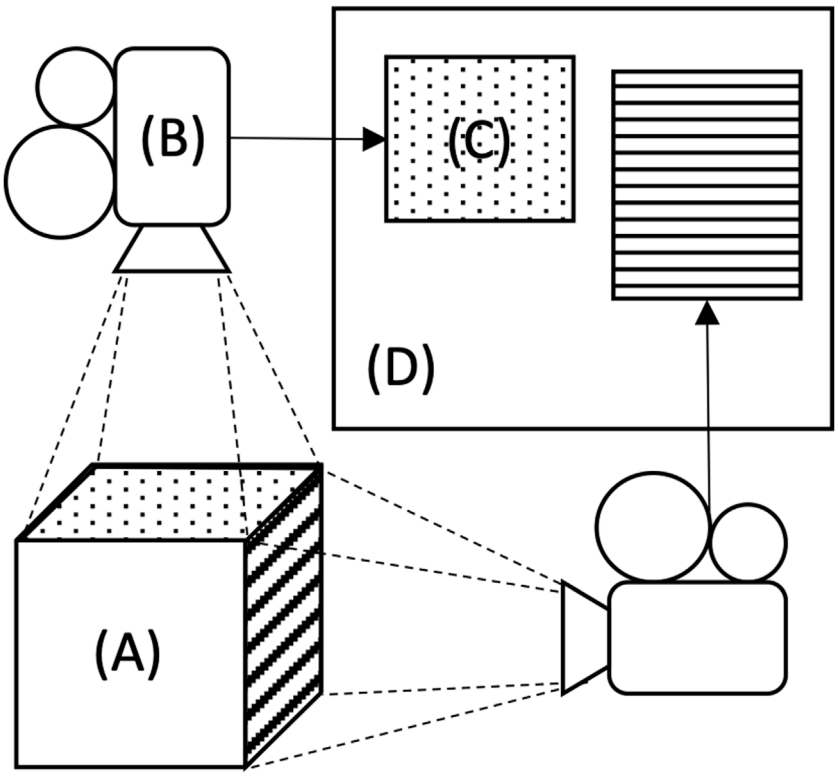

FIGURE 1 | Illustration of conversion from 3D representation of volumetric frame to 2D frame representation.

on support of multiple atlases and the usage of the common atlas data, the reader is referred to (Boyce et al., 2021).

Encoded 2D frame representations and corresponding atlas data are referred to as *V3C components*. A V3C component containing video-encoded 2D frame representations of the volumetric frames is referred to as the *V3C video component* while the V3C component containing the atlas data of the volumetric frames is referred to as the *V3C atlas component*.

Typically, V3C video components represent occupancy, geometry, or attribute information. The occupancy component provides information on which pixels in the other V3C video components are valid and should be used in the 3D reconstruction process. Geometry information, sometimes referred to as the depth map, together with patch information provided through the atlas data indicate the position of the reconstructed voxels in 3D space, while attribute information provides additional properties of the voxels, such as texture color, transparency, or other material information. An example of volumetric video frame projected to occupancy, geometry and attribute components is shown in **Figure 2**. Providing occupancy information as a separate V3C component facilitates the use of padding algorithms in the geometry component to improve the coding performance. While it is required to explicitly signal the occupancy component in the bitstream in the case of V-PCC applications, MIV supports embedding occupancy data within the geometry channel by setting explicit depth-thresholds.

To accommodate devices with limited decoding capabilities, V3C defines a packed V3C video component that combines occupancy, geometry, and attribute information. Packed video components can be particularly useful for catering to platforms where the ability to run multiple parallel video decoding instances is limited. Because the V3C video components are stored in the same video frame, there is also no need for implementing external synchronization between multiple video streams. For more information on packed V3C video components the reader is referred to (Santamaria et al., 2021).

The V-PCC (ISO/IEC 23090-5, 2021) application of V3C specifies additional V3C video components that can carry extra maps of geometry and attribute data. Additional geometry and attribute maps allow to store overlapping projected points, i.e., when a virtual camera would project more than one voxel into the same pixel location in a patch. This creates denser surfaces than using a single map in the reconstruction process and can improve the visual quality of the reconstructed object. In addition to additional maps, V3C specifies a concept of auxiliary data, which can be used to store raw voxels that may have not been projected to the 2D frame representations by the virtual cameras, due to occlusion for example.

The V3C atlas component is encoded using a V3C atlas encoder that creates an atlas bitstream, as shown in **Figure 3**. The atlas encoder minimizes the stored metadata by identifying temporal correspondences between the descriptions of the

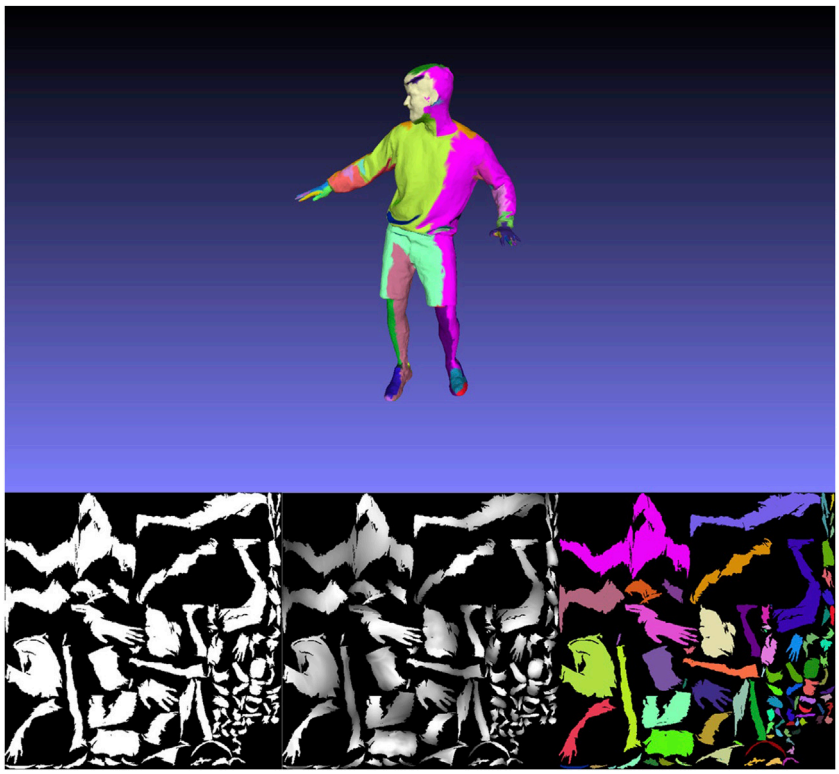

FIGURE 2 | An example of projection of volumetric video frame (top) to 2D representations of occupancy (bottom left), geometry (bottom center), and attribute (bottom right).

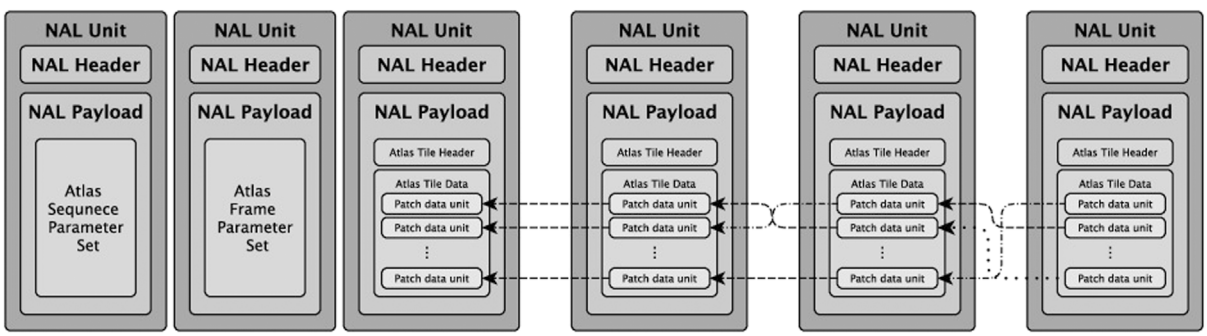

FIGURE 3 | Atlas NAL structure.

patches and storing only the residual information. The high-level syntax of the atlas bitstream uses the concept of Network Abstraction Layer (NAL) units. Like video coding layer (VCL) NAL units in traditional video codecs, such as High Efficiency Video Coding (HEVC), atlas NAL units provide the ability to map the atlas coding layer (ACL) data, which represent the coded atlas data, onto various transport layers. Similar to how a single VCL NAL unit contains one slice of data representing a sub-section of the coded image, a single ACL NAL unit contains one tile of atlas data describing a sub-region of the corresponding coded V3C video components. The concept of tiling facilitates access alignment between atlas data and video data. For example, an atlas tile may be aligned with video slices in HEVC or subpictures in VVC. This enables more efficient parallelization, spatial random access, and partial delivery of volumetric content. The atlas bitstream may also contain non-ACL NAL units which, as the term suggests, contain information not represented in the coded atlas data such as parameter sets or Supplemental Enhancement Information (SEI) messages.

To ensure that a decoder can properly interpret different V3C components, the ISO/IEC 23090-5 specification defines a *V3C bitstream* format. The V3C bitstream encapsulates encoded V3C

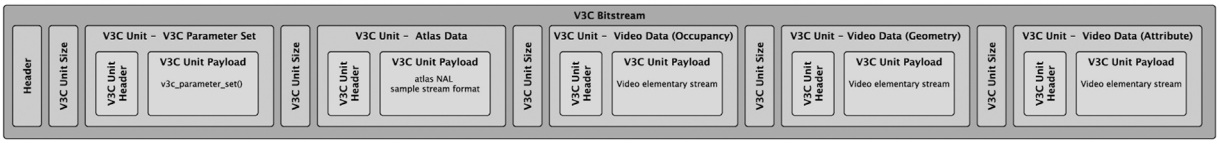

FIGURE 4 | V3C sample stream structure.

video components and V3C atlas components in *V3C units* as illustrated in **Figure 4**. Each V3C unit consists of a V3C unit header and a V3C unit payload pair. The V3C unit header contains information such as atlas ID, component type, map index, attribute index, and a flag indicating if auxiliary data are present. The component type indicates whether the payload contains atlas, geometry, occupancy, or attribute information. V3C bitstreams also contain at least one V3C unit which carries V3C parameter set (VPS) information.

The V3C parameter set structure provides a mechanism for conveying essential data to the decoding process. It indicates the profile, level, and tier of the V3C bitstream, which describe the requirements on the decoder capabilities needed to decode the V3C bitstreams. A V3C profile is composed of three profile components: toolset, codec group, and reconstruction. The toolset component specifies a subset of algorithmic features that a V3C decoder shall support. The codec group component indicates the 2D codec used to encode V3C video components. The reconstruction component indicates a recommended reconstruction method to be used at the decoder side. A level of a tier specifies a set of limits on the values that may be taken by the syntax elements of V3C bitstream, and consequently imply information about the maximum possible memory requirements. More detailed information about the profiles, tiers, and levels defined for V3C can be found in (ISO/IEC 23090-5, 2021).

2.2 ISOBMFF

ISO/IEC 14496-12 (ISO/IEC 14496-12, 2020) specifies an ISO base media file format, commonly referred to as ISOBMFF, that contains timing, structural, and media information for timed sequences of media data such as audio, video, or timed text. The specification also considers storage aspects related to non-timed media such as static images. There are a number of derived specifications that either extend or provide restrictions on the base media file format to define application-specific file formats. The most notable of these are: 1) ISO/IEC 14496-15 (ISO/IEC 14496-15, 2019), which provides carriage of network abstraction layer (NAL) unit structured video codecs such as Advanced Video Coding (AVC) (ISO/IEC 14496-10, 2008), High Efficiency Video Coding (HEVC) (ISO/IEC 23008-2, 2013), and Versatile Video Coding (VVC) (ISO/IEC 23090-3, 2021); 2) a storage format for still images and image sequences, such as exposure stacks defined in High Efficiency Image File Format (HEIF) (ISO/IEC 23008-12, 2017); and 3) ISO/IEC 23090-2 (ISO/IEC 23090-2, 2019), also known as the Omnidirectional Media Format (OMAF), which describes formats and features for omnidirectional videos and still images.

In an ISOBMFF file, data are stored in a series of objects called *boxes*. These boxes may also be nested, where one box contains one or more other boxes as part of its payload. Each box consists of a header, that includes the size of the box in bytes and a four-character code (4CC) field that identifies the box type, i.e. payload format of the box. Commonly, an ISOBMFF file starts with *file type* box, identified by the 4CC 'ftyp', that indicates the specification to which the file complies.

Timed data in an ISOBMFF file is divided into samples which are described by a logical concept called a *track*. A *track* box, identified by the 4CC 'trak', is stored within a *movie* box, 'moov', and contains number of boxes that provide information on where to find the samples in the file and how to interpret them. The samples themselves are stored in a *media data* box identified by the 4CC 'mdat'. Each track contains at least one *sample entry* within a *sample description* box, 'stsd'. A *sample entry* describes the coding and encapsulation format used in the samples of the track. Sample entries are also identified by a 4CC and may contain more boxes for further configurations. One sample entry type defined by the ISOBMFF specification is the *restricted video sample entry* identified by the 4CC 'resv'. This sample entry is used when the decoded sample of a track requires postprocessing before it is passed to a rendering application. The required postprocessing operations are identified by a scheme type signaled within the restricted sample entry itself. **Figure 5** provides a high-level overview of the most important boxes in an ISOBMFF file that includes timed data.

Non-timed data in ISOBMFF is represented by one sample which is described by a logical concept called an *item*. Items are described by a number of boxes that are stored in the *meta* box, 'meta'. Similar to a sample entry in the case of timed data, each item has an *item info entry* that is stored in an *item information* box, 'iinf'. In contrast to timed samples, non-timed sample data may either be stored in the *media data* box, or in an *item data* box, 'idat', that is stored in the *meta* box. **Figure 6** provides a high-level overview of the most important boxes in an ISOBMFF container that stores non-timed data. For more information on storage of items, the reader is referred to (Hannuksela et al., 2015).

2.3 DASH

Dynamic adaptive streaming over HTTP (DASH) is a set of MPEG standards that enable HTTP-based adaptive streaming applications where the media content is captured and stored on a web server and delivered to client players over HTTP. To enable dynamic adaptation in a streaming session, the media content is encoded into several alternatives where each alternative is divided into several segments. The main standard in the MPEG-DASH

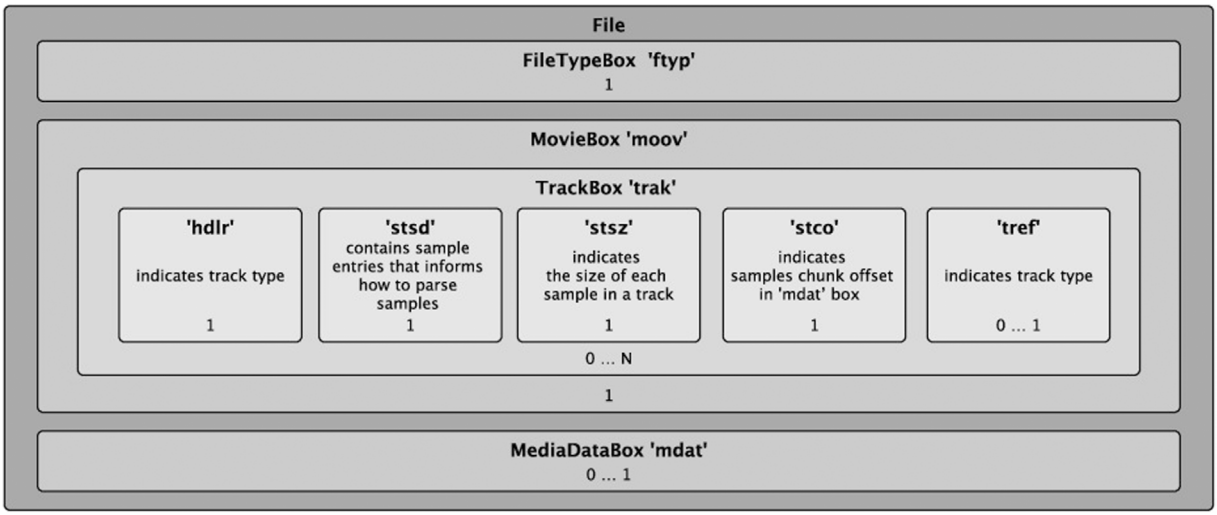

FIGURE 5 | An overview of ISOBMFF file structure for timed data.

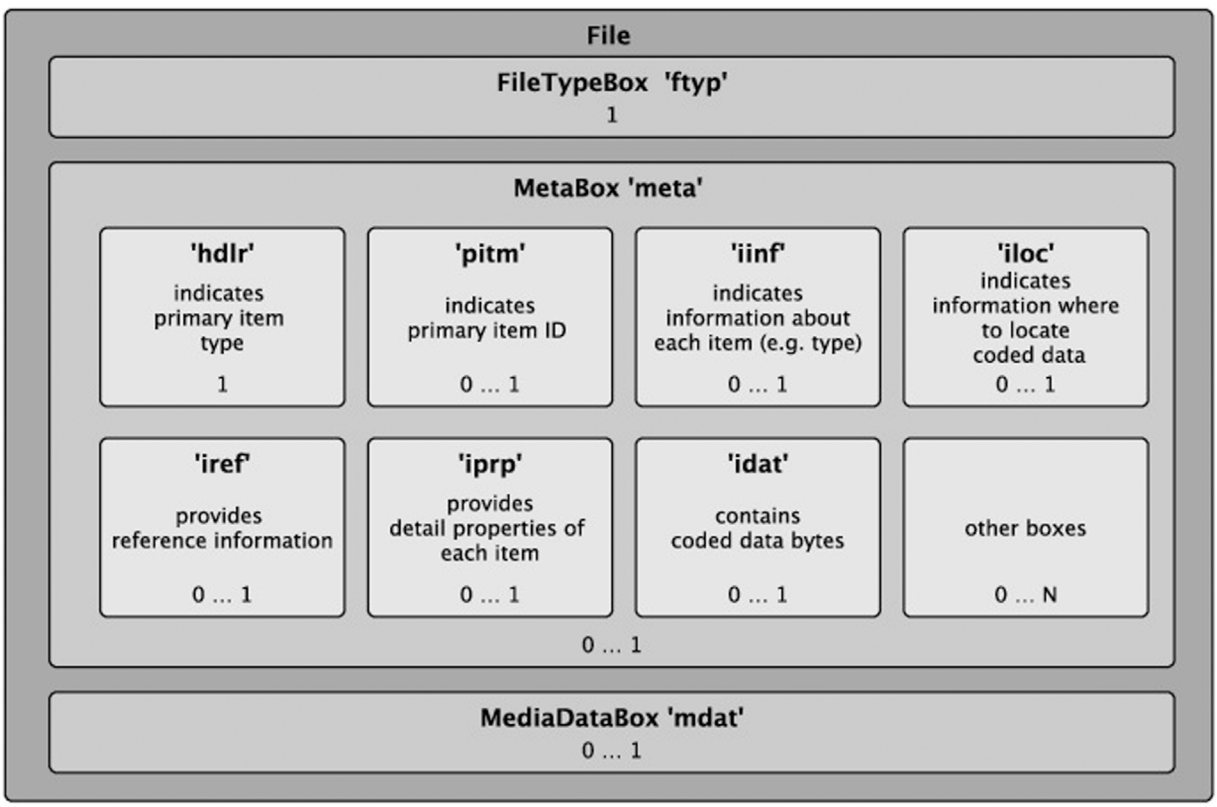

FIGURE 6 | An overview of ISOBMFF file structure for non-timed data.

family is ISO/IEC 23009-1 (ISO/IEC 23009-1, 2019), which specifies the format for a media presentation description (MPD), also known as the manifest, and the segment formats for the streamed content.

The MPD is an Extensible Markup Language (XML) file that describes the streamable content, its components, and the location of all alternative representations. The streamable content is structured into one or more Adaptation Sets, each containing one or more

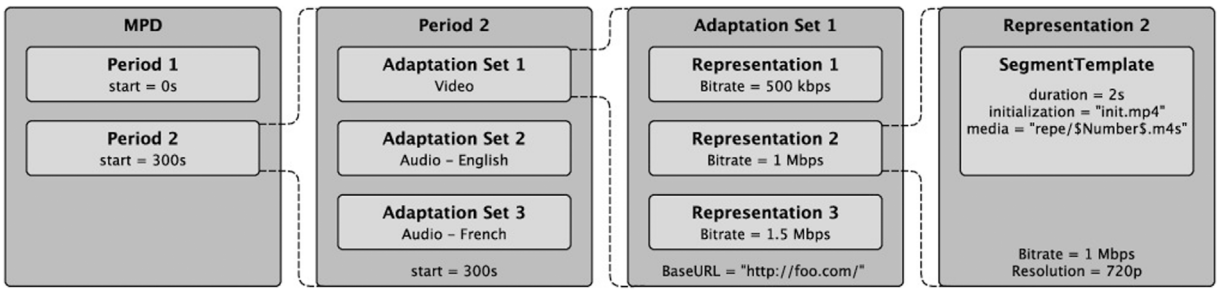

FIGURE 7 | An example of MPD file structure.

Representations. When ISOBMFF is used as an ingest format, a Representation corresponds to a track, whereas an Adaptation Set corresponds to a group of alternative tracks that represent the same content. Each Representation in the Adaptation Set is encoded at different parameters such as bitrate, frame rate or resolution. This enables the DASH player to identify and start playback of the media content and switch between Representations as necessary to adapt to changing network conditions or buffer status, and to change Adaptation Sets to respond to user input. **Figure 7** provides an overview how MPD, Periods, Adaptation Sets and Representation corresponds to each other.

The MPEG-DASH (ISO/IEC 23009-1, 2019) specification defines essential and supplemental property descriptor elements for describing additional characteristics of the Representations or Adaptation Sets. It also defines a Preselection element which enables the content author to define a combination of Adaptation Sets that form a specific experience and can be selected for joint decoding and rendering. When playing a Preselection, a player may select one Representation from each of its constituent Adaptation Sets. Two types of Adaptation Sets are differentiated within a Preselection: a Main Adaptation Set and partial Adaptation Sets. A Representation from the Main Adaptation Set is essential for the playback of the Preselection, while Representations from partial Adaptation Sets are only consumable in combination with the Main Adaptation Set.

There are three basic types of segments in DASH: an Initialization segment, a Media segment, and an Index segment. Initialization segments are used for bootstrapping the media decoding and playback. Media segments contain the coded media data. Index segments provide a directory to the Media segments for a more fine-grained access. An MPD contains either a template for deriving a uniform resource locator (URL) for each segment or a list of specific segment URLs. Players use the URLs (or byte ranges of the URLs) of the selected segments when requesting the content over HTTP.

For more information on DASH, the reader is referred to (Sodagar, 2011).

3 STORAGE OF V3C CONTENT IN ISOBMFF

The ISO/IEC 23090-10 (ISO/IEC 23090-10, 2021) specification is derived from ISO/IEC 14496-12 (ISO/IEC 14496-12, 2020) and specifies how boxes defined in ISO/IEC 14496-12 should be used for storing V3C-coded content. It also defines new boxes required to store a V3C bitstream in an ISOBMFF container. The specification introduces three methods for storing V3C-coded content in ISOBMFF: single-track storage, multi-track storage, and non-timed storage. Single-track storage is intended for enabling the encapsulation of a V3C bitstream directly in ISOBMFF and as such provides limited flexibility and file format level functionality. It was introduced to enable direct encapsulation and ease sharing of data between content production and pre-processing entities. Multi-track storage is indented for commercial deployment scenarios. Multi-track storage encapsulates each V3C component of the V3C bitstream into its own ISOBMFF track. It provides more features for example the ability to directly feed the decoders of the V3C components without de-multiplexing the V3C bitstream. Lastly, the non-timed storage mode enables the storage of static V3C objects, which may be used for example as thumbnails for volumetric video tracks. The three storage modes utilize several common boxes introduced in ISO/IEC 23090-10, such us the *V3C decoder configuration* box, the *V3C unit header* box, the *V3C atlas parameter set sample group description entry*, and the *object switch alternatives entity-to-group* box. The summary of all 4CC values in the document can be found in **Table 1** along with the source specification, where the reader can find further information.

The *V3C decoder configuration* box, 'v3cC', contains a *V3C decoder configuration record* syntax structure which provides information required to: 1) properly interpret the V3C components by storing the V3C parameter set (VPS); 2) initialize an atlas decoder using atlas parameter sets and SEI messages; and 3) understand the track's sample format. The current version of the specification allows to store only one VPS per *V3C decoder configuration* box while the number of atlas parameter sets and SEI messages stored in the sample entry is dependent on the sample entry type.

The *V3C unit header* box, 'vunt', is used to store a four-byte V3C unit header, as defined in ISO/IEC 23090-5 (ISO/IEC 23090-5, 2021). The box allows to properly identify different V3C components and map them to the active VPS. The *V3C unit header* box can be stored in the sample entry of the V3C atlas track as well as in *scheme information box*, 'schi', of V3C video component tracks. Furthermore, the *V3C unit header* box can be

TABLE 1 | Summary of 4CC values.

4CC Value	Type	Description	Source Specification
frma	box	Original format box	ISO/IEC 14496-12
ftyp	box	File-type box	ISO/IEC 14496-12
hdlr	box	Handler reference box	ISO/IEC 14496-12
idat	box	Item data box	ISO/IEC 14496-12
iinf	box	Item information box	ISO/IEC 14496-12
infe	box	Item info entry	ISO/IEC 14496-12
iprp	box	Item properties box	ISO/IEC 14496-12
iref	box	Item reference box	ISO/IEC 14496-12
mdat	box	Media data box	ISO/IEC 14496-12
meta	box	Metadata box	ISO/IEC 14496-12
mmvi	box	Multimap video box	ISO/IEC 23090-10
moov	box	Movie box	ISO/IEC 14496-12
pitm	box	Primary item box	ISO/IEC 14496-12
potg	box	Playout track group box	ISO/IEC 23090-10
resv	sample entry	Restricted video sample entry	ISO/IEC 14496-12
rinf	box	Restricted scheme information box	ISO/IEC 14496-12
schi	box	Scheme information box	ISO/IEC 14496-12
schm	box	Scheme type box	ISO/IEC 14496-12
stsd	box	Sample description box	ISO/IEC 14496-12
swpc	box	Object switch alternatives entity to group box	ISO/IEC 23090-10
tkhd	box	Track header box	ISO/IEC 14496-12
trak	box	Track box	ISO/IEC 14496-12
trgr	box	Track group box	ISO/IEC 14496-12
vaps	sample group	V3C atlas parameter set sample group description entry	ISO/IEC 23090-10
v3a1	sample entry/ item	Atlas sample entry containing all atlas parameter sets when multiple atlases are present, also valid item type for non-timed encapsulation	ISO/IEC 23090-10
v3ag	sample entry	Atlas sample entry with atlas parameter sets potentially in samples when multiple atlases are present	ISO/IEC 23090-10
v3c1	sample entry/ item	Multi-track sample entry containing all parameter sets, when only one atlas is present, also valid item type for non-timed encapsulation	ISO/IEC 23090-10
v3cb	sample entry/ item	Base track sample entry when multiple atlases are present, also valid item type for non-timed encapsulation	ISO/IEC 23090-10
v3cC	box	V3C decoder configuration box or V3C decoder configuration property for non-timed media	ISO/IEC 23090-10
v3cg	sample entry	Multi-track sample entry with parameter sets potentially in samples, when only one atlas is present	ISO/IEC 23090-10
v3cs	reference	Track or item reference to V3C atlas track or V3C atlas item	ISO/IEC 23090-10
v3ct	reference	Track or item reference to V3C atlas tile track or V3C atlas tile item	ISO/IEC 23090-10
v3e1	sample entry	Single-track sample entry containing all parameter sets	ISO/IEC 23090-10
v3eg	sample entry	Single-track sample entry with parameter sets potentially in samples	ISO/IEC 23090-10
v3mt	brand	Basic multi-track encapsulation brand	ISO/IEC 23090-10
v3mp	brand	Advanced multi-track encapsulation brand	ISO/IEC 23090-10
v3nt	brand	Non-timed encapsulation brand	ISO/IEC 23090-10
v3st	brand	Single-track encapsulation brand	ISO/IEC 23090-10
v3t1	sample entry/ item	Atlas tile track sample entry, also valid item type for non-timed encapsulation	ISO/IEC 23090-10
v3tC	box	V3C atlas tile configuration box	ISO/IEC 23090-10
v3va	reference	Track or item reference to attribute V3C video component or attribute V3C component item	ISO/IEC 23090-10
v3vg	reference	Track or item reference to geometry V3C video component or geometry V3C component item	ISO/IEC 23090-10
v3vo	reference	Track or item reference to occupancy V3C video component or occupancy V3C component item	ISO/IEC 23090-10
v3vp	reference	Track or item reference to packed V3C video component or packed V3C component item	ISO/IEC 23090-10
Volv	handler	Handler type for volumetric visual media	ISO/IEC 14496-12
Vunt	box	V3C unit header box	ISO/IEC 23090-10
Vutp	box	V3C unit header property	ISO/IEC 23090-10

used for accessing subsegments of volumetric data, e.g., based on the atlas ID which is present in V3C unit header.

The *V3C atlas parameter set sample group description entry*, 'vaps', is intended for reducing the overhead of temporal random access in V3C atlas tracks. It allows grouping track samples with corresponding common parameter sets and SEI messages and storing these parameter sets and SEI messages in the sample group description without having to repeat them in each sample.

The *object switch alternatives entity to group* box, 'swpc', is used to group V3C atlas tracks and items based on a logical context. The box indicates multiple alternative representations of a V3C-encoded content that represent the same entity, where only one should be played at a given time. For example, an ISOBMFF file may contain two encoded V3C bitstreams of a model where in one bitstream the model is wearing a red dress and in another bitstream the dress is blue. When the tracks

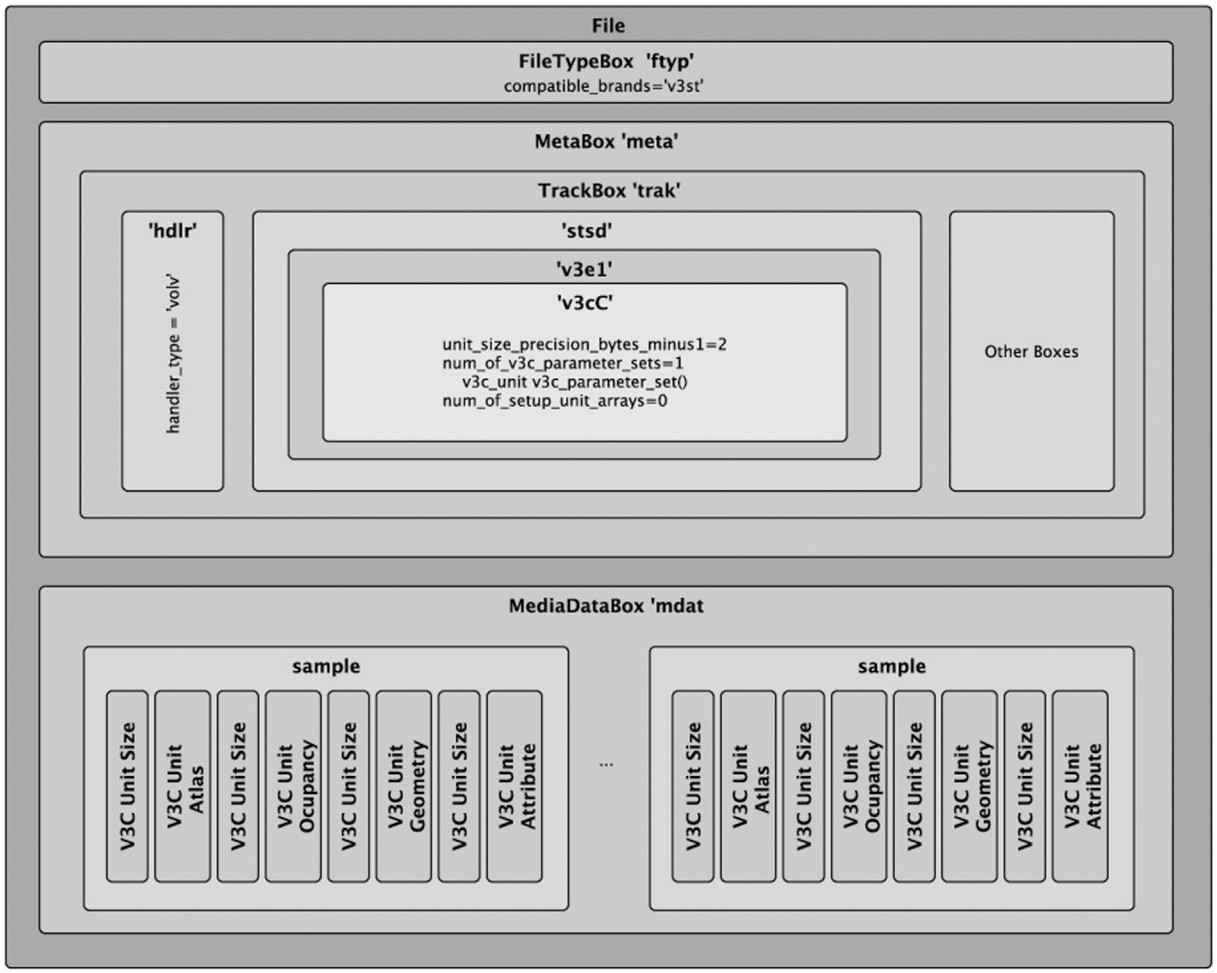

FIGURE 8 | An example of V3C single-track encapsulation.

carrying those two bitstreams are part of an object switch alternative group, the file reader understands that only one should be played at any given time and the user can choose which one should be displayed.

To signal to a parser which storage mode is used and what type of functionality needs to be supported to play the file, ISO/IEC 23090-10 defines four ISOBMFF brands. The single-track encapsulation mode is identified by the brand 'v3st'. Multi-track encapsulation is identified by the 'v3mt' and 'v3mp' brands. The 'v3mt' brand informs the parser that the file contains V3C content stored using a basic multi-track storage mode, while the brand 'v3mp' indicates a multi-track storage mode with additional features present, such as spatial partial access or recommended viewports. These additional features are not described in this article and the interested readers are referred to clauses 9 and 10 of the ISO/IEC 23090-10 specification for more information. Lastly, the non-timed encapsulation mode is identified by the brand 'v3nt'.

3.1 Single-Track Encapsulation

A single-track encapsulation mode represents the V3C bitstream in ISOBMFF as one track, V3C bitstream track. V3C bitstream track is identified by a sample entry with type 'v3e1' or 'v3eg'. This encapsulation mode is intended for direct ISOBMFF encapsulation without any additional pre-processing or de-multiplexing of the ingested V3C bitstream. Each sample in a V3C bitstream track contains one or more V3C units that belong to the same temporal instance, i.e., share the same composition and presentation time. V3C units in a track's sample are stored in the same order they appear in the V3C bitstream. Each V3C unit in a sample is preceded by a number of bytes indicating the size of the following V3C unit. The number of bytes used to indicate the size is provided in *V3C configuration* box.

The simplicity of single-track encapsulation mode has its drawbacks. The most evident one is the lack of partial access. As all V3C units are stored in a single sample, a client is not able to select whether to consume all V3C components or only a portion of them.

Figure 8 provides a high-level overview of the boxes present in an ISOBMFF file containing V3C bitstream in a single-track encapsulation mode and the relationships between them. As shown in the figure, the only information extracted from the V3C bitstream and stored in *V3C configuration* box are the V3C

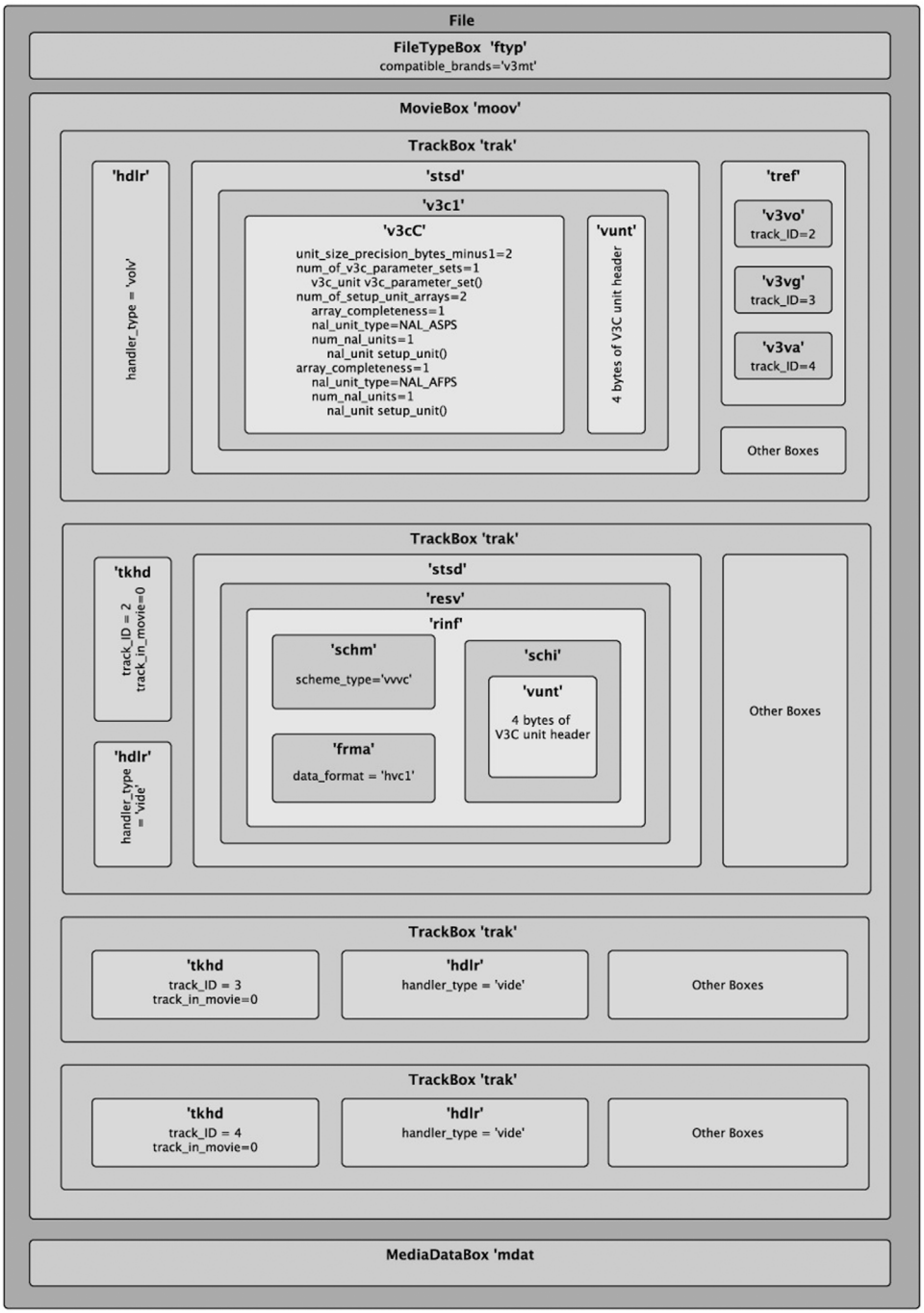

FIGURE 9 | An example of V3C bitstream stored in multi-track mode.

parameter set and *unit_size_precision_bytes_minus1*, used to indicate size of the V3C units. The rest of the V3C bitstream, i.e., the encoded video data and atlas data, is stored as is in the *media data* box.

3.2 Multi-Track Encapsulation

A multi-track encapsulation mode stores the V3C bitstream in the ISOBMFF file as several tracks, where each track represents either part of or a complete V3C component. This mode is well suited for streaming workflows where independent encoders can work in parallel and the resulting bitstreams can be stored into an ISOBMFF-compliant file or set of files as separate tracks. It also allows easy extraction and direct processing of each V3C component by their respective decoder without the need to reconstruct the V3C bitstream. This is done by exposing to the file parser the required information for identifying

individual V3C components, or parts thereof, so that only the relevant information is provided to the client application.

Tracks in the multi-track encapsulation mode are divided into V3C atlas tracks, V3C atlas tile tracks, and V3C video component tracks based on the type of V3C component they encapsulate. An ISOBMFF file that is compliant to the multi-track encapsulation mode must contain at least one V3C atlas track, which is the entry point to the content. The number of other V3C atlas tracks, V3C atlas tile tracks, and V3C video component tracks in the file depends on the nature of the ingested V3C bitstream as well the needs of the application the ISOBMFF file is created for. The samples of tracks originating from the same V3C bitstream are time-aligned, which means that all samples from different tracks contributing to the same volumetric frame must have the same composition time.

To bundle multiple tracks into logical entities, the multi-track encapsulation mode uses track references to establish relationships between the different track types. This allows parsers to quickly select appropriate tracks from the file based on the client application's preferences.

Figure 9 provides a high-level overview of the boxes present in an ISOBMFF file that utilizes the multi-track encapsulation mode to store a V3C bitstream.

3.2.1 V3C Atlas Tracks

A V3C atlas track describes the data of V3C atlas component and is identified by a sample entry of type 'v3c1', 'v3cg', 'v3cb', 'v3a1', or 'v3ag'. Each V3C atlas track sample entry contains a *V3C configuration* box as well as a *V3C unit header* box.

A V3C atlas track with a sample entry of type 'v3c1' or 'v3cg' is used when V3C bitstream contains only a single atlas, a typical V-PCC use case. The sample entry type 'v3c1' indicates that all atlas parameter sets are stored in the sample entry and application does not need to consider atlas parameter sets appearing in the samples. The sample entry type 'v3cg' indicates that atlas parameter sets may be stored in the sample entry or in the samples of the track.

V3C atlas tracks with a sample entry of type 'v3cb', 'v3a1', and 'v3ag' are used when V3C bitstream contains more than one atlas, a typical MIV use case. In this scenario ISOBMFF file contains at least one V3C atlas track with a sample entry of type 'v3cb' and one or more V3C atlas tracks with a sample entry of type 'v3a1' or 'v3ag'. V3C atlas track with a sample entry of type 'v3cb' is the entry point for parsing the ISOBMFF file and consists of only the common atlas data of the ingested V3C bitstream. It also references other V3C atlas tracks with the sample entry types 'v3a1' or 'v3ag' by using a track reference of the type 'v3cs'. V3C atlas tracks with a sample entry of type 'v3a1' or 'v3ag' describe the atlas data of the ingested V3C bitstream. The sample entry of type 'v3a1' indicates that all atlas parameter sets are stored in the sample entry while the sample entry of type 'v3cg' indicates that atlas parameter sets may be stored in the sample entry or in the samples.

V3C atlas track with a sample entry 'v3c1', 'v3cg', 'v3a1' and 'v3ag' can refer to V3C video component tracks using a reference type which indicates the component type of the referenced track:

- 'v3vo' indicates that the referenced video track contains an occupancy V3C component,
- 'v3vg' indicates that the referenced video track contains a geometry V3C component,
- 'v3va' indicates that the referenced video track contains an attribute V3C component, and
- 'v3vp' indicates that the referenced video track contains a packed V3C component.

Each sample of a V3C atlas track corresponds to a coded atlas access unit or a coded common atlas access unit. Each sample contains atlas NAL units of data encapsulated in the V3C sample stream NAL unit format, as defined in ISO/IEC 23090-5 (2021), which means that each atlas NAL unit is preceded by a size value indicating the number of bytes for the NAL unit. The information on how many bytes are used to indicate the size value is provided in the *V3C configuration* box.

3.2.2 V3C Atlas Tile Track

A V3C atlas tile track describes a portion of V3C atlas component, and it is identified by a sample entry of type 'v3t1'. The sample entry of V3C atlas tile tracks contains a *V3C atlas tile configuration* box 'v3tC', which identifies the tile IDs for the tiles that the track describes. V3C atlas tile tracks are always associated with another V3C atlas track in the ISOBMFF file. The V3C atlas track refers to V3C atlas tile tracks using a track reference of type 'v3ct'. When V3C atlas tile tracks are present in the ISOBMFF file, they refer to the V3C video component tracks that contain corresponding partitions of the V3C video components. This creates a hierarchy of references, as illustrated in **Figure 10**, where the parser identifies the entry point to the content by finding the V3C atlas track, then follows track references to V3C atlas tile tracks, which in turn refer to the V3C video component tracks for the indicated tiles. V3C atlas tile tracks can be used in combination with multiple atlases, in which case the track references create an additional level of hierarchy between V3C atlas tracks as discussed in the previous section.

Samples of V3C atlas tile tracks are stored in the same sample format as V3C atlas tracks. The difference is that the sample data must only contain ACL NAL units particular to the tiles indicated in the *V3C atlas tile configuration* box.

The tiling concept in a V3C atlas allows dividing an object represented by the volumetric frame, either spatially or otherwise, into independent sub-regions. When atlas tiles are aligned with video codecs partitioning concepts, such as motion constrained slices in HEVC or sub-pictures in VVC, V3C atlas tile tracks provide an efficient mechanism for implementing partial access functionality.

3.2.3 V3C Video Component Tracks

A V3C video component track describes the encoded data of a V3C video component. It is a restricted video track, not intended to be displayed by standard 2D video players, identified by a sample entry of type 'resv'. The sample entry of a V3C video component track includes a *restricted scheme information* box, 'rinf', containing at least three other boxes: 1) a *scheme type* box, 'schm', indicating a 'vvvc' scheme; 2) a *scheme information* box,

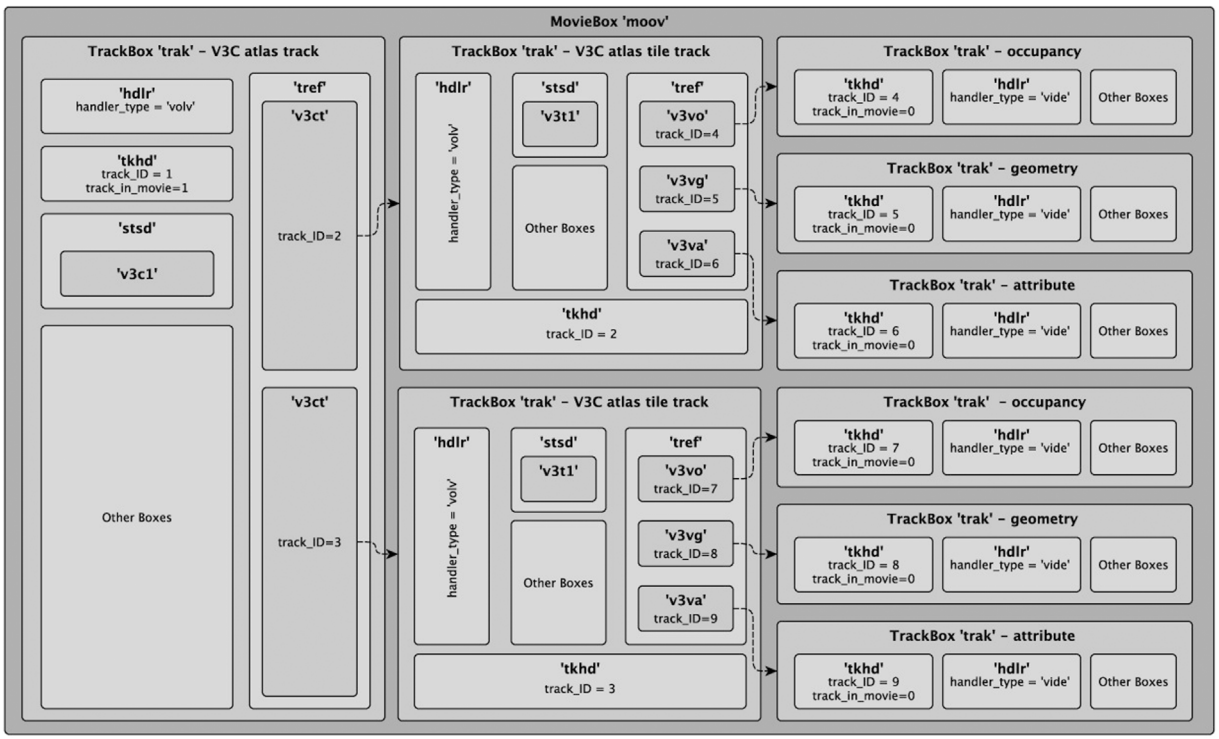

FIGURE 10 | Illustration of references with atlas tile tracks.

'schi', with a *V3C unit header* box identifying the type of the V3C video component; and 3) an *original format* box, 'frma', storing the original sample entry type of the video track. The samples of a V3C video component track are stored in the format indicated by the original sample entry type in the *original format* box.

A V3C video component in a V3C bitstream may temporally multiplex multiple maps. To indicate this information at the ISOBMFF-level and ensure proper timestamp assignment to samples by a file parser, a V3C video track can include a *multimap video* box, 'mmvi', in the *scheme information* box.

3.2.4 Track Alternatives and Grouping

Like traditional video, a volumetric video can be encoded in different alternative representations. In ISOBMFF, the alternate_group field is used to indicate the tracks that are part of the same alternative group. The ISO/IEC 23090-10 specification defines two levels of adaptation using track alternatives. In the first level, V3C atlas tracks which have the same value for the alternate_group field in the *track header* box, 'tkhd', are considered as alternatives to each other. In the second level, V3C video component tracks can have alternative representations, similar to traditional video tracks. Utilization of V3C atlas track or V3C video component track alternatives allows applications to scale the quality of playback according to the users' preferences. For example, when the client device's resolution does not benefit from a high-resolution video component or when the target device has limited video decoding resources.

However, combining different alternatives of V3C video component tracks arbitrarily may sometimes result in poor rendering quality. This can be the case when combining a low-resolution occupancy V3C video component track with a high-resolution geometry V3C video component track for example. To avoid such undesired combinations of track alternatives, an ISOBMFF container creator can indicate the recommended combinations of tracks that can be played together to the client by using a *playout track group* box, 'potg', stored in the *track group* box, 'trgr', of the tracks.

3.3 Non-Timed Encapsulation

A non-timed encapsulation mode represents the V3C bitstream in ISOBMFF as items, each item represents part or whole of one V3C component. This mode is well suited, for example, to store V3C bitstream encoded with a still picture profile that has less strict decoding requirements and limited coding tool support.

Items originating from a V3C bitstream are described in a *meta* box with the *handler reference* box, 'hdlr', indicating handler type 'volv'. V3C atlas component is described by a V3C atlas item with item type 'v3c1', 'v3cb' or 'v3a1'. V3C atlas item with item type 'v3c1' is used when V3C bitstream contains only a single atlas while V3C atlas items with item type 'v3cb' and 'v3a1' are used when V3C bitstream contains more than one atlas. When V3C atlas item with item type 'v3cb' is present, one or more V3C atlas items with item type 'v3a1' are present as well. In that case V3C atlas item with item type 'v3cb' is the entry point for parsing the ISOBMFF file and contains only the common atlas data of the

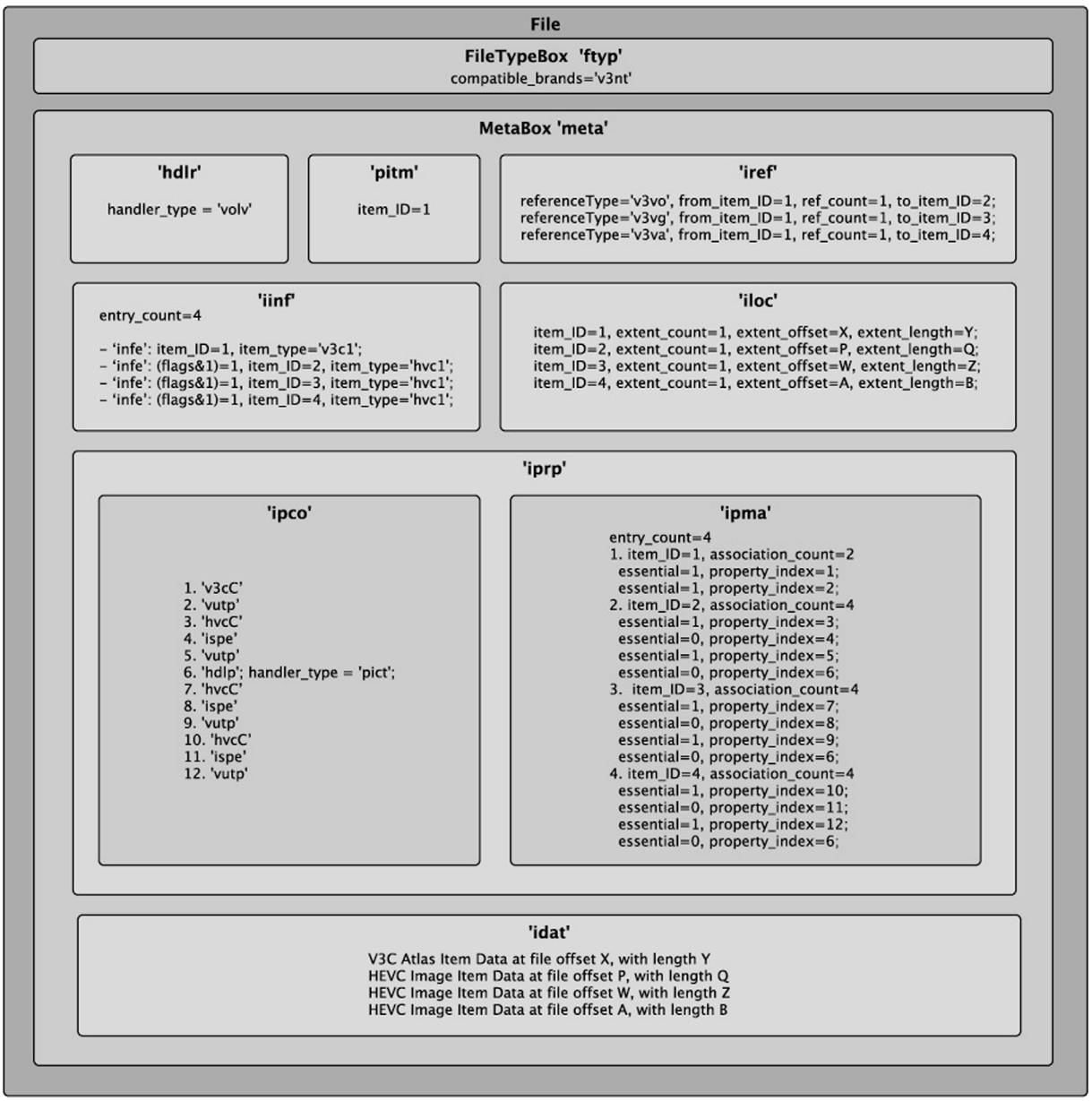

FIGURE 11 | An example of V3C item with three V3C video components stored as HEVC items.

ingested V3C bitstream. V3C video components are described by V3C video component items with the item type corresponding to the still image codec, e.g., in case of still HEVC encoded image an item type of 'hvc1'is used.

V3C atlas item with type 'v3c1' or 'v3cb' is always indicated as the primary item in a *primary item* box, 'pitm'. When the primary item is of type 'v3c1', it is linked with items describing V3C video components, using the entries in the *item reference* box, 'iref', with same reference types as defined for multi-track encapsulation mode ('v3vo', 'v3vg', 'v3va' and 'v3vp'). When the primary item is of type 'v3cb', it is linked with other V3C atlas items of type 'v3a1' using reference type 'v3cs'.

Using the *item properties* box, 'iprp', each of the items constituting to the V3C bitstream can be defined. The V3C atlas item is associated at least with *V3C configuration* property, 'v3cC', and *V3C unit header* property, 'vutp', both properties has same syntax and function as *V3C configuration* box and *V3C unit header* box in the multi-track encapsulation mode.

The items corresponding to V3C video components should be associated with at least a *V3C unit header* property. The *V3C unit header* property identifies the V3C video component and maps it to VPS information. To inform regular file readers that the items describing V3C video components should not

TABLE 2 | Examples of V3C MPD.

```
<?xml version="1.0" encoding="UTF-8"?>
<MPD>
<Period duration="PT10S">

    <!—Example 1: Single track V3C AdaptationSet -->
    <AdaptationSet
        mimeType="video/mp4" codecs="v3e1.L2.0.0.1, resv.vvvc.avc1.4D401E" frameRate="30">
        <SegmentList>
            <Initialization sourceURL="seg-m-init.mp4"/>
        </SegmentList>
        <Representation bandwidth="512000">
            <BaseURL>vpcc-512k.mp4</BaseURL>
        </Representation>
        <Representation bandwidth="1024000">
            <BaseURL>vpcc-1024k.mp4</BaseURL>
        </Representation>
    </AdaptationSet>

    <!-- Example 2 -->
    <!-- Main V3C AdaptationSet -->
    <AdaptationSet id="1" codecs="v3c1">
           <EssentialProperty schemeIdUri="urn:mpeg:dash:preselection:2016" />
           <Representation>
               <SegmentList>
                       <Initialization sourceURL="v3c_init.mp4" />
                       ...
               </SegmentList>
           </Representation>
    </AdaptationSet>

    <!-- Occupancy -->
    <AdaptationSet id="2" mimeType="video/mp4" codecs="resv.vvvc.hvc1">
           <EssentialProperty schemeIdUri="urn:mpeg:dash:preselection:2016" />
           <EssentialProperty schemeIdUri="urn:mpeg:mpegI:v3c:2020:videoComponent">
               <v3c:videoComponent type="occp" />
           </EssentialProperty>
           <Representation>
               ...
           </Representation>
    </AdaptationSet>

    <!-- Geometry -->
    <AdaptationSet id="4" mimeType="video/mp4" codecs="resv.vvvc.hvc1">
           <EssentialProperty schemeIdUri="urn:mpeg:dash:preselection:2016" />
           <EssentialProperty schemeIdUri="urn:mpeg:mpegI:v3c:2020:videoComponent">
               <v3c:videoComponent type="geom" />
           </EssentialProperty>
           <Representation>
               ...
           </Representation>
    </AdaptationSet>

    <!-- Attribute -->
    <AdaptationSet id="6" mimeType="video/mp4" codecs="resv.vvvc.hvc1">
           <EssentialProperty schemeIdUri="urn:mpeg:dash:preselection:2016" />
           <EssentialProperty schemeIdUri="urn:mpeg:mpegI:v3c:2020:videoComponent">
               <v3c:videoComponent type="attr" attribbute_type="0" attribute_index="0" />
           </EssentialProperty>
           <Representation>
               ...
           </Representation>
    </AdaptationSet>

    <!-- Preselections -->
    <Preselection id="1" tag="1" preselectionComponents="1 2 4 6" codecs="v3c1">
           <!—V3C Descriptor -->
           <SupplementalProperty schemeIdUri="urn:mpeg:mpegI:v3c:2020:v3c" vId="1" />
    </Preselection>

</Period>
</MPD>
```

be displayed, the items are marked as hidden by setting the (*flags* and 1) equal to 1 in their respective *item info entry*, 'infe'.

Figure 11 shows an example of V3C bitstream stored according to non-timed encapsulation mode. The coded data, V3C atlas item data and HEVC image item data, are stored in an *item data* box, but could be also stored in *media data* box.

Finally, like V3C atlas tile tracks in the multi-track encapsulation mode, a V3C atlas tile items with a 'v3t1' item type can be created in the non-timed encapsulation mode. To indicate the relationship between a V3C atlas item and V3C atlas tile item, item reference with type 'v3ct' is used.

4 STREAMING

The ISO/IEC 23090-10 specification supports delivering V3C content using MPEG-DASH (ISO/IEC 23009-1, 2019). The standard defines how to signal V3C content in the Media Presentation Description (MPD) for both the single-track and multi-track encapsulation modes as described in **Sections 3.1** and

Section 3.2, respectively, and defines restrictions on the DASH segments generated for the content.

4.1 Single-Track Mode

The single-track mode enables streaming of V3C ISOBMFF files where V3C content is stored using single-track encapsulation mode. In this mode, the V3C content is represented with a single Adaptation Set in the MPD with one or more Representations. The only constraint on the Representations of this Adaptation Set is that the codec used for encoding a given V3C video component must be identical across all Representations. There is no requirement, however, that all the V3C video components in one Representation must be encoded using the same codec.

In the single-track mode, the Initialization Segment must contain the V3C parameter set and the component codec mapping SEI messages, defined in the ISO/IEC 23090-5, in the sample entry of the main V3C atlas track. In addition, the first sample of a Media Segment must have a Stream Access Point (SAP) of type 1 or 2, as defined in the ISO/IEC 23009-1 (2019).

Example one in **Table 2** contains an example of an MPD with V3C media content signaled in the single-track mode. In the example, the sample entry of the V3C atlas track is 'v3e1' and the content is encoded according to main tier (L) with level 2. The profile used during encoding was V-PCC basic toolset with Rec1 reconstruction. V3C video components are encoded with AVC, with additional information signaled through resv.vvvc.avc1.4D401E. The example describes a single V3C content in a single Adaptation Set as two alternative Representations.

4.2 Multi-Track Mode

The multi-track mode provides more flexibility over the single-track mode by enabling adaptation across several dimensions as each V3C component is represented by its own Adaptation Set. ISO/IEC 23090-10 distinguishes four types of Adaptation Sets: 1) Main Adaptation Set 2) Atlas Adaptation Set, 3) Atlas Tile Adaptation Set, and 4) Video Component Adaptation Set.

An Adaptation Set representing a V3C atlas track with a 'v3c1', 'v3cg', or 'v3cb' sample entry type serves as the Main Adaptation Set for accessing the V3C content described in the MPD. In case the V3C content contains more than one atlas, the Adaptation Set representing the V3C atlas track with sample entry type 'v3cb' is the Main Adaptation Set and the Adaptation Sets describing the remaining atlases, with either sample entry type 'v3a1' or 'v3ag', are referred to as Atlas Adaptation Sets. The relationship between those Adaptation Sets is signaled through the *dependencyId* attribute, which is set in the Representations of each Atlas Adaptation Set to an ID of a Representation in the Main Adaptation Set. More granular adaptations can be exposed in the MPD by employing the tiling concept of the V3C codec. In that situation each V3C atlas tile track with sample entry type 'v3t1' is described by an Atlas Tile Adaptation Set.

Adaptation Sets describing V3C video components, including maps or auxiliary data, are referred to as Video Component Adaptation Sets. For the Video Component Adaptation Sets, the codecs attribute is set based on the respective codec used for encoding the video component and takes the form 'resv.vvvc.xxxx', where xxxx corresponds to the 4CC of the video codec signaled in the *restricted scheme information* box ('rinf') of the sample entry for the corresponding V3C video component track. To indicate the relationship between Video Component Adaptation Sets and Main, Atlas, and Atlas Tile Adaptation Sets, the concept of Preselections, described in **Section 4.2.3**, is used.

By separating the V3C video component bitstreams into multiple Adaptation Sets, a client can prioritize or completely drop some components or maps when making adaptation decisions. Additionally, V3C video component representations can be encoded using different video codecs or different bitrates to allow adaptive bitrate streaming. In such scenario, ISO/IEC 23090-10 mandates that the Representations with different video codecs should be signaled as a separate Video Component Adaptation Sets in the MPD. Additionally, to indicate support of seamless switching between Representations across different Video Component Adaptation Sets, a Supplemental Property descriptor with the *schemeIdUri* attribute set to "urn:mpeg:dash:adaptation-set-switching:2016" should be present in the switchable Video Component Adaptation Sets. The *value* attribute of this descriptor provides a comma-separated list of the Video Component Adaptation Set IDs to which a player can switch.

4.2.1 Segment Constraints

The multi-track mode defines number of constrains on the segments of the Adaptation Sets describing V3C components. The Main Adaptation Set shall contain a single Initialization Segment that includes all parameter sets needed to initialize the V3C decoder. A Media Segment in a Representation of the Main Adaptation Set or Atlas Adaptation Sets contains one or more track fragments of the V3C atlas track, while a Media Segment in a Representation of a Video Component Adaptation Set contains one or more track fragments of the corresponding V3C video component track.

An ISOBMFF file conforming to the multi-track encapsulation mode can be generated by concatenating the Initialization Segment of the Main Adaptation Set with the subsegments from the Representations of the Main Adaptation Set, Atlas Adaptation Set, and Video Component Adaptation Sets associated with the Atlas Adaptation Set. The resulting file can be parsed by a conforming player.

4.2.2 V3C and V3C Video Component Descriptors

Adaptation Sets in a multi-track MPD describe different V3C components. To ensure that a player can identify the type and characteristics of a given Adaptation Set. ISO/IEC 23090-10 defines two new types of descriptors: a V3C descriptor and a V3C Video Component descriptor.

The V3C descriptor is provided as a Supplemental Property descriptor and it is identified by a *schemeIdUri* with value equal to "urn:mpeg:mpegI:v3c:2020:v3c". The V3C descriptor provides three attributes *vId, atlas_id,* and *tile_ids*. The presence of each depends on where the descriptor is used; Main Adaptation Set, an Atlas Tile Adaptation Set, a V3C Preselection, or an Atlas Tile Preselection. When a V3C

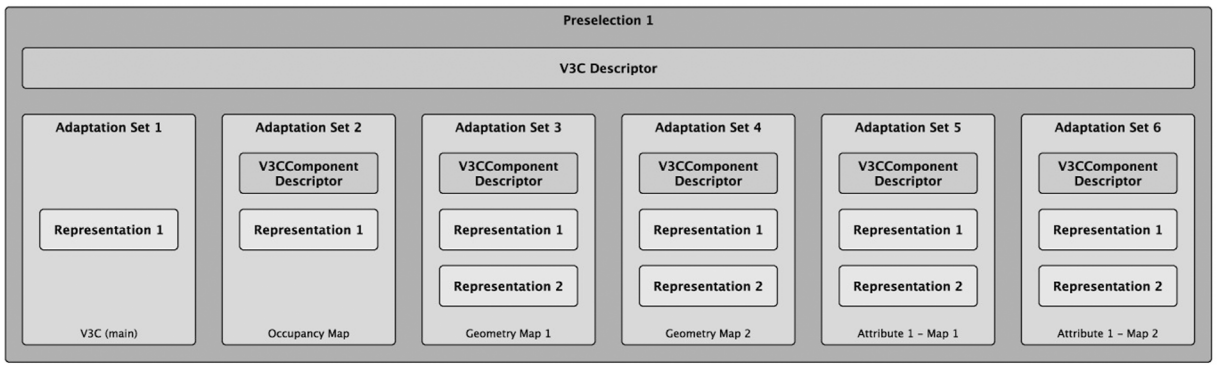

FIGURE 12 | An example of a V3C Preselection.

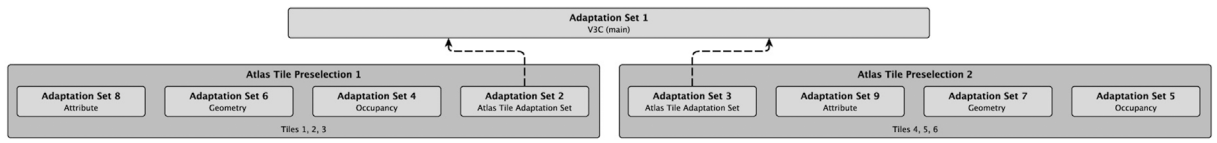

FIGURE 13 | An example of V3C Atlas Tile Preselections.

descriptor is used in the Main Adaptation Set or a V3C Preselection, it includes only the *vId* attribute indicating a unique ID of the V3C content in the MPD. When a V3C descriptor is used in an Atlas Adaptation Set or an Atlas Preselection, it also includes the *atlas_id* attribute signaling the atlas ID of the associated Adaptation Set or Preselection. Lastly, when a V3C descriptor is used in an Atlas Tile Adaptation Set or an Atlas Tile Preselection, in addition to the other attributes, it includes the *tile_ids* attribute describing the subset of the tile IDs that the Adaptation Set or Preselection contains.

The V3C Video Component descriptor is provided as an Essential Property descriptor and it is identified by a *schemeIdUri* with a value equal to "urn:mpeg:mpegI:v3c:2020: videoComponent". The V3C Video Component descriptor is present in every Video Component Adaptation Set and provides several attributes that indicate the type and properties of the V3C video component. For example, if the descriptor indicates that the Adaptation Set contains an attribute V3C video component, additional attributes, such as *attribute_type*, *attribute_index,* and *map_index*, are provided to identify the attribute V3C video component among other attribute V3C video components.

4.2.3 V3C, Atlas, and Atlas Tile Preselections

The concept of Preselections, which is defined in ISO/IEC 23009-1 (Sodagar, 2011), is used to group all Adaptation Sets describing V3C components into a single decodable user experience. ISO/IEC 23090-10 defines three types of Preselections: V3C Preselections, Atlas Preselections, and Atlas Tile Preselections.

A V3C Preselection contains an Atlas Adaptation Set as the main Adaptation Set of the Preselection and the Video Component Adaptation Sets associated with the atlas as partial Adaptation Sets. This preselection is used for the most basic use case where the V3C content is composed of a single atlas and no tiling division is exposed on ISOBMFF level. This use case is presented in **Figure 12**.

V3C Atlas Preselections are used when the V3C content is composed of more than one atlas. This enables a player to identify the group of Adaptation Sets in an MPD that are associated with exactly one atlas.

A V3C Atlas Tile Preselection is used when the V3C content contains an atlas with multiple tiles, which are represented as separate Atlas Tile Adaptation Sets. This preselection groups each Atlas Tile Adaptation Set and associated Video Component Adaptation Sets. It should be noted that a V3C Atlas Tile Preselection is not decodable on its own and depends on the Adaptation Set of the atlas to which the tiles belong. This dependency is indicated in the Representations of the Atlas Tile Adaptation Set using the *dependencyId* attribute which should be set to the ID of a Representation in the corresponding Atlas Adaptation Set. An example of a usage of V3C Atlas Tile Preselection is shown on **Figure 13**. In the example, a V3C content is represented by two V3C atlas tile tracks described by two Atlas Tile Adaptation Sets and number of Video Component Adaptation Sets. Two V3C Atlas Tile Preselections are used, each one groups Atlas Tile Adaptation Set and associated Video Component Adaptation Sets.

4.2.4 MPD Example

Example two in **Table 2** contains an example of an MPD with V3C content signaled in the multi-track mode. In the example, a V3C content is represented by four Adaptation Sets and a V3C

Preselection. The Main Adaptation Set is indicated by setting the *codecs* attribute to the 4CC of the V3C atlas track it represents and contains the Initialization Segment for the V3C content. Each Video Component Adaptation Sets is identified by a V3C Video Component descriptor with the *type* set "occp", "geom", or "attr". Video Component Adaptation Sets, along with the Main Adaptation Set, are grouped using a V3C Preselection. To indicate that these Adaptation Sets are referenced in a Preselection, a Preselection descriptor without the *value* attribute is signaled in each Adaptation Set. The V3C Preselection includes a V3C descriptor that indicates the mandatory *vId* attribute. For brevity, the Representations of the Adaptation Sets were omitted in the example.

5 CONCLUSIONS AND FUTURE WORK

This article has provided an overview of the technologies specified in the ISO/IEC 23090-10 standard, which together with ISO/IEC 23090-5 and ISO/IEC 23090-12 form a V3C family of standards developed by MPEG that enable compression, storage, and delivery of volumetric videos. The standards exploit the ubiquity of traditional 2D video coding and delivery technologies, aiming to minimize additional investments required to enable distribution of volumetric video to the masses.

The article has described the different encapsulation modes supported by the ISO/IEC 23090-10 standard and how the different components in an encoded V3C bitstream can be stored in an ISOBMFF file. Various DASH extensions, specified in ISO/IEC 23090-10, related to delivery of V3C media over a network were introduced to the reader.

Combining the toolsets and features, provided by the V3C family of standards, with the traditional 2D video coding tools, such as motion-constrained tiles in HEVC or sub-pictures in VVC, enables the deployment of advanced and efficient delivery systems that cater the content based on network conditions and end user preferences. These systems present opportunities for further research, including, but not limited to, optimal tiling strategies and bitrate adaptation for tile-based streaming.

Going forward, the future will show how successful the V3C family of standards will become. But some positive early market indications are already visible. The Sistema Brasileiro de Televisão Digital (SBTVD) industry forum has chosen V3C as a candidate for their VR codec (SBTVD), while the Virtual Reality Industry Forum (VRIF) has recently issued the first version of industry guidelines for volumetric video streaming (VR Industry Forum, 2021), based fully on V3C. Moreover, discussion on real-time streaming aspects of V3C have been initialized in IETF (IETF Audio/Video Transport Core Maintenance (avtcore) Working Group, 2021), with the goal of standardizing Real-time Transport Protocol (RTP) payload format for V3C bitstream.

While the first family of V3C standards is finalized, the work on standardization of volumetric media coding and delivery still continues. MPEG is working on new applications such as mesh-coding as well as on improvements to the base specifications. To demonstrate the validity of the technologies in the V3C family of standards, conformance and reference software is continuously being developed within MPEG.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Boyce, J. M., Dore, R., Dziembowski, A., Fleureau, J., Jung, J., Kroon, B., et al. (2021). MPEG Immersive Video Coding Standard. *Proc. IEEE* 109 (9), 1521–1536. doi:10.1109/jproc.2021.3062590

Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T., and Tabatabai, A. (2020). An Overview of Ongoing point Cloud Compression Standardization Activities: Video-Based (V-PCC) and Geometry-Based (G-PCC). *APSIPA Trans. Signal Inf. Process.* 9 (13), e13. doi:10.1017/atsip.2020.12

Hannuksela, M. M., Lainema, J., and Malamal Vadakital, V. K. (2015). The High Efficiency Image File Format Standard [Standards in a Nutshell]. *IEEE Signal. Process. Mag.* 32 (4), 150–156. doi:10.1109/msp.2015.2419292

IETF Audio/Video Transport Core Maintenance (avtcore) Working Group (2021). Interim Meeting Agenda. [Online]. Available: https://datatracker.ietf.org/doc/agenda-interim-2022-avtcore-01-avtcore-01/ (Accessed February 15, 2022).

Iso/Iec 14496-10 (2008). *Information Technology — Coding of Audio-Visual Objects — Part 10: Advanced Video Coding.*

Iso/Iec 14496-12 (2020). *Information Technology — Coding of Audio-Visual Objects — Part 12: ISO Base media File Format.*

Iso/Iec 14496-15 (2019). *Information Technology — Coding of Audio-Visual Objects — Part 15: Carriage of Network Abstraction Layer (NAL) Unit Structured Video in the ISO Base media File Format.*

Iso/Iec 23008-1(2017). *Information Technology — High Efficiency Coding and media Delivery in Heterogeneous Environments — Part 1: MPEG media Transport (MMT).*

Iso/Iec 23008-12(2017). *Information Technology — High Efficiency Coding and media Delivery in Heterogeneous Environments — Part 12: Image File Format.*

Iso/Iec 23008-2 (2013). *Information Technology — High Efficiency Coding and media Delivery in Heterogeneous Environments — Part 2: High Efficiency Video Coding.*

Iso/Iec 23009-1(2019). *Information Technology – Dynamic Adaptive Streaming over HTTP (DASH) – Part 1: Media Presentation Description and Segment Formats.*

Iso/Iec 23090-10 (2021). *Information Technology — Coded Representation of Immersive media — Part 10: Carriage of Visual Volumetric Video-Based Coding Data.*

Iso/Iec 23090-12 (2021). *Information Technology — Coded Representation of Immersive media — Part 12: MPEG Immersive Video.*

Iso/Iec 23090-2 (2019). *Information Technology — Coded Representation of Immersive media — Part 2: Omnidirectional media Format.*

Iso/Iec 23090-3 (2021). *Information Technology — Coded Representation of Immersive media — Part 3: Versatile Video Coding.*

Iso/Iec 23090-5 (2021). *Information Technology — Coded Representation of Immersive media — Part 5: Visual Volumetric Video-Based Coding (V3C) and Video-Based point Cloud Compression (V-PCC).*

Kongsilp, S., and Dailey, M. N. (2017). Motion Parallax from Head Movement Enhances Stereoscopic Displays by Improving Presence and Decreasing Visual Fatigue. *Displays* 49, 72–79. doi:10.1016/j.displa.2017.07.001

Recommendation H.261 (1988). *International Telecommunication Union - Telecommunication Standardization Sector (ITU-T), H.261: Video Codec for Audiovisual Services at P X 384 Kbit/s*. Geneva, Switzerland: ITU.

Santamaria, M., Malamal Vadakital, V. K., Kondrad, L., Hallapuro, A., and Hannuksela, M. M. (2021). "Coding of Volumetric Content with MIV Using VVC Subpictures," in Proceeding of the IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP), Tampere, Finland, Oct. 2021 (IEEE). doi:10.1109/mmsp53017.2021.9733465

SBTVD. TV 3.0 Project. [Online]. Available: https://forumsbtvd.org.br/tv3_0/ (Accessed December 3, 2021).

Schwarz, S., Preda, M., Baroncini, V., Budagavi, M., Cesar, P., Chou, P. A., et al. (2019). Emerging MPEG Standards for Point Cloud Compression. *IEEE J. Emerg. Sel. Top. Circuits Syst.* 9 (1), 133–148. doi:10.1109/jetcas.2018.2885981

Sodagar, I. (2011). The MPEG-DASH Standard for Multimedia Streaming over the Internet. *IEEE MultiMedia* 18 (4), 62–67. doi:10.1109/mmul.2011.71

VR Industry Forum (2021). *Volumetric Video Guidelines, Public Draft v0.95*.

Conflict of Interest: Authors LI, LK, and SS were employed by the company Nokia Solutions and Networks GmbH & Co. KG. Author AH was employed by the company InterDigital Canada Ltée.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ilola, Kondrad, Schwarz and Hamza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.