DRAFT INTERNATIONAL STANDARD

ISO/IEC DIS 23090-10

ISO/IEC JTC **1**/SC **29** Secretariat: **JISC**

Voting begins on:
2020-10-05

Voting terminates on:
2020-12-28

Information technology — Coded representation of immersive media —

Part 10:
Carriage of visual volumetric video-based coding data

ICS: 35.040.40

This document is circulated as received from the committee secretariat.

Reference number
ISO/IEC DIS 23090-10:2020(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/IEC JTC 1 Information technology, Subcommittee SC 29, Coding of audio, picture, multimedia and hypermedia information.

A list of all parts in the ISO 23090 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document addresses the storage of visual volumetric video-based coding data in files based on ISO/IEC 14496-12, reusing existing tools for storage of video-coded components. Another important aspect considered by this specification is supporting flexible extraction of component streams at delivery and/or decoding time.

Information technology — Coded representation of immersive media— Part 10: Carriage of Visual Volumetric Video-based Coding Data

1 Scope

This document specifies carriage of coded media representations which comply with visual volumetric video-based coding and video-based point cloud compression specification, ISO/IEC 23090-5 [V3C].

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

[14496-12]	ISO/IEC 14496-12, Information technology — Coding of audio-visual objects — Part 12: ISO base media file format
[14496-15]	ISO/IEC 14496-15, Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO based media file format
[V3C]	ISO/IEC 23090-5, Information technology — Coded representation of immersive media — Part 5: Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC)
[DASH]	ISO/IEC 23009-1, Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats
[XMLS]	W3C Recommendation, XML schema part 1: Structures
[XMLD]	W3C Recommendation, XML schema part 2: Datatypes
[RFC6381]	IETF RFC 6381, The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types
[23008-12]	ISO/IEC 23008-12, Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 12: Image File Format

3 Terms and definitions

In this specification, the terms which are defined in ISO/IEC 23090-5 [V3C] apply. Additionally, the following terms are defined.

3.1
V3C content
volumetric media that is encoded using ISO/IEC 23090-5.

3.2
shall
a term used to express mandatory requirements for conformance to this Specification

3.3
V3C track
volumetric visual track which carries the atlas data V3C component of the V3C bitstream

3.4
video-coded V3C component track
video track which carries 2D video encoded data for any of the occupancy map, geometry, or attribute component video bitstreams of the V3C bitstream

3.5
viewport
planar viewing region used to project portion of the 3D scene to the position and orientation of a virtual camera

3.6
viewing orientation
triple of azimuth, elevation, and tilt angle characterizing the orientation that a user is consuming the audio-visual content; in case of image or video, characterizing the orientation of the viewport

3.7
viewing position
triple of x, y, z characterizing the position in the global reference coordinate system of a user who is consuming the audio-visual content; in case of image or video, characterizing the position of the viewport

[Ed(EA): The terms viewport, viewport orientation and viewing position are added to terms and definitions as agreed in MPEG#130. However, they are not referenced in the specification text.]

4 Abbreviations

For the purposes of this International Standard, the following abbreviations apply:

2D	Two-Dimensional
CVS	Coded V3C Sequence
DASH	MPEG Dynamic Adaptive Streaming over HTTP (specified in ISO/IEC 23009-1)
IRAP	Intra Random Access Point
ISOBMFF	ISO Base Media File Format (specified in ISO/IEC 14496-12)
PCC	Point Cloud Compression
SEI	Supplemental Enhancement Information
V3C	Visual Volumetric Video-based Coding
VPS	V3C Parameter Set
V-PCC	Video-based PCC

5 Volumetric media

5.1 Volumetric visual media

5.1.1 General

A volumetric visual track shall be identified by the volumetric visual media handler type `'volv'` in the `HandlerBox` of the `MediaBox`, as defined in ISO/IEC 14496-12 [14496-12], and by a volumetric visual media header as defined in this specification.

Multiple volumetric visual tracks may be present in the file.

5.1.2 Volumetric visual media header

5.1.2.1 Definition

Box Type: `'vvhd'`
Container: `MediaInformationBox`
Mandatory: Yes
Quantity: Exactly one

Volumetric visual tracks shall use a `VolumetricVisualMediaHeaderBox` in the `MediaInformationBox` as defined in ISO/IEC 14496-12 [14496-12].

5.1.2.2 Syntax

```
aligned(8) class VolumetricVisualMediaHeaderBox
   extends FullBox('vvhd', version = 0, 1) {
}
```

[Ed (EA): It should be noted that volumetric media header, handler and volumetric sample entry should be specified in ISOBMFF and referenced by this specification]

5.1.2.3 Semantics

`version` is an integer that specifies the version of this box

5.1.3 Volumetric visual sample entry

5.1.3.1 Definition

Volumetric visual tracks shall use a `VolumetricVisualSampleEntry`

5.1.3.2 Syntax

```
class VolumetricVisualSampleEntry(codingname)
   extends SampleEntry (codingname){
   unsigned int(8)[32] compressorname;
   // other boxes from derived specifications
}
```

5.1.3.3 Semantics

`compressorname` is a name, for informative purposes. It is formatted in a fixed 32-byte field, with the first byte set to the number of bytes to be displayed, followed by that number of bytes of displayable data encoded using UTF-8, and then padding to complete 32 bytes total (including the size byte). The field may be set to 0.

5.1.4 Volumetric visual samples

The format of a volumetric visual sample is defined by the coding system.

6 Carriage of Visual Volumetric Video-based Coding Data

6.1 General

A V3C bitstream, containing coded V3C sequence (CVS), is composed of V3C units carrying V3C parameter set (VPS) data, a coded atlas bitstream, a 2D video encoded occupancy map bitstream, a 2D video encoded geometry bitstream, and zero or more 2D video encoded attribute bitstreams.

6.2 Common boxes and data structures

6.2.1 V3C decoder configuration record

6.2.1.1 Definition

The V3C decoder configuration record provides V3C bitstream's decoding specific information (i.e. parameter sets and SEI messages) for further configuration and initialization of the V3C decoder.

In this version of the specification,

- `configurationVersion` shall be equal to 1.
- `numOfV3CParameterSets` shall be equal to 1.

The `v3cParameterSet` array shall include a single `v3c_parameter_set()`, as defined in ISO/IEC 23090-5 [V3C].

If the value of `configurationVersion` is not supported by a reader, then it shall not attempt to decode this configuration record or the bitstreams to which it applies.

The `setupUnit` arrays shall include atlas sub-bitstream parameter sets that are constant for the CVS referred to by the sample entry in which the decoder configuration record is present as well as atlas sub-bitstream SEI messages.

6.2.1.2 Syntax


```
aligned(8) class V3CDecoderConfigurationRecord {
   unsigned int(8) configurationVersion = 1;
   unsigned int(2) lengthSizeMinusOne;
   bit(1) reserved = 1;
   unsigned int(5) numOfV3CParameterSets;
   for (i=0; i < numOfV3CParameterSets; i++) {
      unsigned int(16) V3CParameterSetLength;
      v3c_unit(V3CParameterSetLength) v3cParameterSet;  // as defined in ISO/IEC
23090-5
   }
   unsigned int(8) numOfSetupUnitArrays;
   for (j=0; j < numOfSetupUnitArrays; j++) {
      bit(1) array_completeness;
      bit(1) reserved = 0;
      unsigned int(6) NAL_unit_type;
      unsigned int(8) numNALUnits;
      for (i=0; i < numNALUnits; i++) {
         unsigned int(16) SetupUnitLength;
         nal_unit(SetupUnitLength) setupUnit;  // as defined in ISO/IEC 23090-5
      }
   }
   // additional fields
}
```

6.2.1.3 Semantics

`configurationVersion` is a version field. Incompatible changes to the record shall be indicated by a change of version number.

`lengthSizeMinusOne` plus 1 indicates the length in bytes of the `NALUnitLength` field in a V3C sample in the stream to which this configuration record applies. For example, a size of one byte is indicated with a value of 0. The value of this field shall be equal to `ssnh_unit_size_precision_bytes_minus1` in `sample_stream_nal_header()` for the atlas substream.

`numOfV3CParameterSets` specifies the number of V3C parameter set units signalled in the decoder configuration record.

`V3CParameterSetLength` indicates the size, in bytes, of the `v3cParameterSet` field.

`v3cParameterSet` is a V3C unit of type `V3C_VPS`, as defined in ISO/IEC 23090-5.

`numOfSetupUnitArrays` indicates the number of arrays of atlas NAL units of the indicated type(s).

`array_completeness` when equal to 1 indicates that all atlas NAL units of the given type are in the following array and none are in the stream; when equal to 0 indicates that additional atlas NAL units of the indicated type may be in the stream; the default and permitted values are constrained by the sample entry name.

`NAL_unit_type` indicates the type of the atlas NAL units in the following array (which shall be all of that type); it takes a value as defined in ISO/IEC 23090-5; it is restricted to take one of the values indicating a `NAL_ASPS`, `NAL_AAPS`, `NAL_AFPS`, `NAL_PREFIX_ESEI`, `NAL_PREFIX_NSEI`, `NAL_SUFFIX_ESEI`, or `NAL_SUFFIX_NSEI` atlas NAL unit.

`numNALUnits` indicates the number of atlas NAL units of the indicated type included in the configuration record for the stream to which this configuration record applies. The SEI array shall only contain SEI messages of a 'declarative' nature, that is, those that provide information about the stream as a whole. An example of such an SEI could be a user-data SEI.

`SetupUnitLength` indicates the size, in bytes, of the `setupUnit` field. The length field includes the size of both the NAL unit header and the NAL unit payload but does not include the length field itself.

`setupUnit` may contain a NAL unit of type `NAL_ASPS`, `NAL_AAPS`, `NAL_AFPS`, `NAL_PREFIX_ESEI`, `NAL_PREFIX_NSEI`, `NAL_SUFFIX_ESEI`, or `NAL_SUFFIX_NSEI`, as defined in ISO/IEC 23090-5. When present in `setupUnit`, `NAL_PREFIX_SEI` or `NAL_SUFFIX_SEI` contains SEI messages of a 'declarative' nature, that is, those that provide information about the stream as a whole. An example of such an SEI could be a user-data SEI.

6.2.2 V3C decoder configuration box

A V3C decoder configuration box includes the a `V3CDecoderConfigurationRecord` as defined in 6.2.1.

6.2.2.1 Syntax

```
class V3CConfigurationBox extends Box('v3cC') {
    V3CDecoderConfigurationRecord() V3CConfig;
}
```

6.2.2.2 Semantics

`V3CDecoderConfigurationRecord` is defined in 6.2.1

6.2.3 V3C unit header box

6.2.3.1 Definition

V3C unit header box is present in a sample entry of a V3C track and in scheme information of all video-coded V3C component tracks. It contains the V3C unit header for the data carried by the respective track.

6.2.3.2 Syntax

```
aligned(8) class V3CUnitHeaderBox
   extends FullBox('vunt', version = 0, 0) {
   v3c_unit_header   unit_header;  // as defined in ISO/IEC 23090-5
}
```


6.2.3.3 Semantics

`unit_header` is a `v3c_unit_header()` as defined in ISO/IEC 23090-5.

6.2.4 V3C atlas parameter set sample group

6.2.4.1 General

The use of `'vaps'` for the `grouping_type` in sample grouping represents the assignment of samples in V3C track to the atlas parameter sets carried in this sample group. When a `SampleToGroupBox` with `grouping_type` equal to `'vaps'` is present, an accompanying `SampleGroupDescriptionBox` with the same grouping type shall be present and contains the ID of the group that the samples belong to.

A V3C track may contain at most one `SampleToGroupBox` with `grouping_type` equal to `'vaps'`.

When the `'v3c1'` or `'v3e1'` sample entry is used in V3C track, NAL units of type NAL_ASPS, NAL_AAPS, NAL_AFPS, NAL_PREFIX_ESEI, NAL_PREFIX_NSEI, NAL_SUFFIX_ESEI, and NAL_SUFFIX_NSEI (as defined in ISO/IEC 23090-5 [V3C]) shall not be present in the samples of this sample group.

When the `'v3cg'` or `'v3eg'` sample entry is used in V3C track, NAL units of type NAL_ASPS, NAL_AAPS, NAL_AFPS, NAL_PREFIX_ESEI, NAL_PREFIX_NSEI, NAL_SUFFIX_ESEI, and NAL_SUFFIX_NSEI may be present in the samples of this sample group.

6.2.4.2 Syntax


```
aligned(8) class V3CAtlasParamSampleGroupDescriptionEntry() extends
SampleGroupDescriptionEntry('vaps') {
    unsigned int(8) numOfSetupUnits;
    for (i=0; i < numOfSetupUnits; i++) {
       unsigned int(16) setupUnitLength;
       nal_unit(setupUnitLength) setupUnit; // as defined in ISO/IEC 23090-5
    }
}
```


6.2.4.3 Semantics

`numOfSetupUnits` specifies the number of setup units signalled in the sample group description.

`setupUnitLength` indicates the size, in bytes, of the `setupUnit` field. The length field includes the size of both the NAL unit header and the NAL unit payload but does not include the length field itself.

`setupUnit` is a NAL unit of type `NAL_ASPS`, `NAL_AAPS`, `NAL_AFPS`, `NAL_PREFIX_ESEI`, `NAL_PREFIX_NSEI`, `NAL_SUFFIX_ESEI`, or `NAL_SUFFIX_NSEI` carrying data associated with this group of samples.

6.3 Single track encapsulation of V3C data

6.3.1 General

A single-track encapsulation of V3C data requires the V3C bitstream to be represented by a single-track declaration.

Single-track encapsulation of V3C data could be utilized in the case of simple ISOBMFF encapsulation of a V3C encoded bitstream. Such a bitstream could be directly stored as a single track without further processing. V3C unit header data structures can be kept in the bitstream as-is. A single-track encapsulated V3C data could be provided to media workflows for further processing (e.g., multi-track file generation, transcoding, DASH segmentation, etc.).

6.3.2 V3C bitstream track

6.3.2.1 V3C bitstream track sample entry

Sample Entry Type: `'v3e1'`, `'v3eg'`
Container: `SampleDescriptionBox`
Mandatory: A `'v3e1'` or `'v3eg'` sample entry is mandatory
Quantity: One or more sample entries may be present

V3C bitstream tracks shall use `VolumetricVisualSampleEntry` with a sample entry type of `'v3e1'` or `'v3eg'`.

A V3C bitstream track sample entry shall contain a `V3CConfigurationBox`, as defined in 6.2.2

Under the '`v3e1`' sample entry, all atlas parameter sets and SEI messages (as defined in ISO/IEC 23090-5 [V3C]) shall be in the `setupUnit` array. Under the `'v3eg'` sample entry, the atlas parameter sets and SEI messages may be present in the `setupUnit` array, or in the V3C bitstream.

An optional `BitRateBox` as defined in [14496-12] may be present in the V3C bitstream sample entry to signal the bit rate information of the V3C bitstream track.

6.3.2.1.1 Syntax

```
class V3CConfigurationBox extends Box('v3cC') {
    V3CDecoderConfigurationRecord() V3CConfig;
}

aligned(8) class V3CBitstreamSampleEntry() extends VolumetricVisualSampleEntry
('v3e1') {
   V3CConfigurationBox config;
}
```


6.3.2.1.2 Semantics

`compressorname` in the base class `VolumetricVisualSampleEntry` indicates the name of the compressor used with the value "`\012V3C Coding`" being recommended; the first byte is a count of the remaining bytes, here represented by `\012`, which (being octal 12) is 10 (decimal), the number of bytes in the rest of the string.

6.3.2.2 V3C bitstream sample format

A V3C bitstream sample shall contain one or more V3C units which belong to the same presentation time, i.e, one V3C access unit. A sample may be self-contained (e.g., a sync sample) or decoding-wise dependent on other samples of V3C bitstream track.

6.3.2.3 V3C bitstream sync sample

A V3C bitstream sync sample shall satisfy all the following conditions:

- It shall be independently decodable.
- None of the samples that come after the sync sample (in decoding order) have any decoding dependency on any sample prior to the sync sample.
- All samples that come after the sync sample (in decoding order) are successfully decodable.

6.3.2.4 V3C bitstream sub-sample

A V3C bitstream sub-sample is a V3C unit which is contained in a V3C bitstream sample.

A V3C bitstream track shall contain one `SubSampleInformationBox` in its `SampleTableBox`, or in the `TrackFragmentBox` of each of its `MovieFragmentBoxes`, which lists the V3C bitstream sub-samples.

The 32-bit unit header of the V3C unit which represents the sub-sample shall be copied to the 32-bit `codec_specific_parameters` field of the sub-sample entry in the `SubSampleInformationBox`. The V3C unit type of each sub-sample shall be identified by parsing the `codec_specific_parameters` field of the sub-sample entry in the `SubSampleInformationBox`.

6.4 Multi-track encapsulation of V3C data

6.4.1 General

The general layout of a multi-track encapsulated V3C data container is shown in Figure 1, where V3C units in a V3C bitstream are mapped to individual tracks within the container file based on their types. There are two types of tracks in a multi-track encapsulated V3C data container: V3C track and V3C component track, defined in clause 3 of this specification.

V3C component tracks are restricted video scheme tracks. In addition, the following conditions shall be satisfied for V3C component tracks:

a) in the sample entry, a new box is inserted which documents the role of the video stream contained in this track, in the V3C system;
b) a track reference is introduced from the V3C track, to the V3C component track, to establish the membership of the V3C component track in the specific volumetric media represented by the V3C track;
c) the track-header flags are set to 0, to indicate that this track does not contribute directly to the overall layout of the movie but contributes to the V3C system.

Tracks belonging to the same CVS are time-aligned. Samples that contribute to the same volumetric media frame across the different video-encoded V3C component tracks and the V3C track shall have the same presentation time. The V3C atlas sub-bitstream parameter sets used for such samples shall have a decoding time equal or prior to the composition time of the volumetric media frame. In addition, all tracks belonging to the same CVS shall have the same implied or explicit edit lists.

Note: Synchronization between the elementary streams in the component tracks are handled by the ISOBMFF track timing structures (`stts`, `ctts`, and `cslg`), or equivalent mechanisms in movie fragments.

Note: The sync samples in the V3C track and V3C component tracks may or may not be time-aligned. In the absence of time-alignment, random access may involve pre-rolling the various tracks from different sync start-times, to enable starting at the desired time. In the case of time-alignment (e.g. required by a V3C profile such as the Basic toolset profile as defined in [V3C]), the sync samples of the V3C track should be considered as the random access points for the V3C content, and random access should be done by only referencing the sync sample information of the V3C track.