FINAL
DRAFT

INTERNATIONAL
STANDARD

**ISO/IEC
FDIS
23090-5**

ISO/IEC JTC **1**/SC **29**

Secretariat: **JISC**

Voting begins on:
2023-08-07

Voting terminates on:
2023-10-02

Information technology — Coded representation of immersive media —

Part 5: Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC)

Technologie de l'information — Représentation codée de média immersifs —

Partie 5: Codage basé sur la vidéo volumétrique (V3C) et compression de nuage de points basée sur la vidéo (V-PCC)

Reference number
ISO/IEC FDIS 23090-5:2023(E)

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

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

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 document 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 or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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. In the IEC, see www.iec.ch/understanding-standards.

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

This second edition cancels and replaces the first edition (ISO/IEC 23090-5:2021), which has been technically revised.

The main changes are as follows:

— Additions needed for ISO/IEC 23090-12 MPEG Immersive video

— Various minor improvements and corrections.

A list of all parts in the ISO/IEC 23090 series can be found on the ISO and IEC websites.

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 and www.iec.ch/national-committees.

Introduction

Advances in 3D capturing and rendering technologies have unleashed a new wave of innovation in Virtual/Augmented/Mixed reality (VR/AR/MR) content creation and communication, of which visual volumetric video is an integral part.

Visual volumetric video, a sequence of visual volumetric frames, if uncompressed, may be represented by a large amount of data, which can be costly in terms of storage and transmission. This has led to the need for a high coding efficiency standard for the compression of visual volumetric data.

Visual volumetric frames can be coded by converting the 3D volumetric information into a collection of 2D images and associated data. The converted 2D images can be coded using widely available video and image coding specifications, such as ISO/IEC 14492-10 and ISO/IEC 23008-2 and the associated data can be coded with mechanisms specified in this document. The coded images and the associated data can then be decoded and used to reconstruct the 3D volumetric information. This document specifies a generic mechanism for visual volumetric video coding, i.e. visual volumetric video-based coding. The generic mechanism may be used by applications targeting volumetric content, such as point clouds, immersive video with depth, mesh representations of visual volumetric frames, etc.

In addition to the generic mechanism of coding volumetric content, this document specifies one of the applications of visual volumetric video-based coding targeting point cloud representations of visual volumetric frames. In a point cloud sequence, each point cloud frame contains a collection of points. Each point has a 3D position, i.e., geometry information, and each point may also be associated with a number of attributes, such as colour, reflectance, surface normal, etc.

Information technology — Coded representation of immersive media —

Part 5: Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC)

1 Scope

This document specifies the syntax, semantics, and decoding for visual volumetric media using video-based coding methods. Furthermore, this document specifies processes that may be needed for reconstruction of visual volumetric media, and may also include additional processes such as post-decoding, pre-reconstruction, post-reconstruction, and adaptation.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitute 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.

ISO/IEC 10646, *Information technology — Universal coded character set (UCS)*

ISO/IEC 14496-10, *Information technology — Coding of audio-visual objects — Part 10: Advanced video coding*

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

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 base media file format*

ISO/IEC 23008-2, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

ISO/IEC/DIS 23090-3, *Information technology — Coded representation of immersive media — Part 3: Versatile video coding*[1)]

ISO/IEC 23091-2, *Coding-independent code points for video signal type identification*

Rec. ITU-T T.35, *Procedure for the allocation of ITU-T defined codes for non standard facilities*

Rec. ITU-T H.271, *Video back-channel messages for conveyance of status information and requests from a video receiver to a video sender*

IEEE 754-2019, *IEEE Standard for Floating-Point Arithmetic*

IETF RFC 1321, *The MD5 Message-Digest Algorithm*

IETF RFC 5646, *Tags for Identifying Languages*

1) Under preparation. Stage at time of publication: ISO/IEC DIS 23090-3:2020.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
3D bounding box
volume defined as a cuboid solid having six rectangular faces placed at right angles

3.2
associated non-ACL NAL unit
non-ACL NAL unit (3.81) that is associated with an *ACL NAL unit* (3.5) for the purpose of decoding or other operations specified by this document

3.3
atlas
collection of 2D bounding boxes and their associated information placed onto a rectangular frame and corresponding to a volume in 3D space on which volumetric data is rendered

3.4
atlas bitstream
sequence of bits that forms the representation of *atlas frames* (3.7) and associated data forming one or more *CASs* (3.35)

3.5
atlas coding layer NAL unit
ACL NAL unit
collective term for coded atlas tile layer *NAL units* (3.79) and the subset of *NAL units* (3.79) that have reserved values of nal_unit_type that are classified as being of type class equal to ACL in this document

3.6
atlas coordinates
two scalars (x, y) with finite precision and dynamic range that indicate the location of an atlas sample relative to the top left corner of an atlas frame, with x and y indicating the horizontal and vertical direction, respectively

3.7
atlas frame
2D rectangular array of atlas samples onto which *patches* (3.88) are projected and additional information related to the *patches* (3.88), corresponding to a *volumetric frame* (3.139)

3.8
atlas frame parameter set
AFPS
syntax structure (3.121) containing *syntax elements* (3.120) that apply to zero or more entire coded *atlas frames* (3.7) as determined by the content of a *syntax element* (3.120) found in each tile header

3.9
atlas sample
position on the rectangular frame onto which *patches* (3.88) that are associated with an *atlas* (3.3) are projected

3.10
atlas sequence
collection of *atlas frames* (3.7)

3.11
atlas sequence parameter set
ASPS
syntax structure (3.121) containing *syntax elements* (3.120) that apply to zero or more entire *coded atlas sequences* (3.34) as determined by the content of a *syntax element* (3.120) found in the AFPS referred to by a syntax element found in each tile header

3.12
atlas sub-bitstream
extracted *sub-bitstream* (3.109) from the *V3C bitstream* (3.129) containing a part of an atlas NAL bitstream

3.13
atlas unit
set of *NAL units* (3.79) that contain all *ACL NAL units* (3.5) of a *coded atlas* (3.33) and their *associated non-ACL NAL units* (3.2)

3.14
attribute
scalar or vector property optionally associated with each point in a *volumetric frame* (3.139) such as colour, reflectance, surface normal, transparency, material ID, etc.

3.15
attribute access unit
collection of *attribute maps* (3.17) and auxiliary attribute frames, if available, for a specific *attribute* (3.14) that correspond to the same time instance

3.16
attribute frame
2D rectangular array created through the aggregation of *patches* (3.88) containing values of a specific *attribute* (3.14)

3.17
attribute map
attribute frame (3.16) containing attribute *patch* (3.88) information projected at a particular depth indicated by the corresponding *geometry map* (3.64)

3.18
auxiliary attribute frame
2D rectangular array that is associated with *RAW patches* (3.104) and *EOM patches* (3.56), and contains values of a specific *attribute* (3.14)

3.19
auxiliary geometry frame
2D rectangular array that is associated with *RAW patches* (3.104) and contains *geometry* (3.61) values

3.20
auxiliary video component
video component (3.126) indicated as being of auxiliary type through an appropriate flag in the *VPS* (3.135) and containing data only associated with *RAW patches* (3.104) and/or *EOM patches* (3.56)

3.21
auxiliary video sub-bitstream
video sub-bitstream (3.128) indicated as being of auxiliary type through an appropriate flag in the *VPS* (3.135) and containing data only associated with *RAW patches* (3.104) and/or *EOM patches* (3.56)

3.22
bitstream
ordered series of bits that forms the coded representation of the data

3.23
byte
sequence of 8 bits, within which, when written or read as a sequence of bit values, the left-most and right-most bits represent the most and least significant bits, respectively

3.24
broken link access access unit
BLA access unit
access unit in which the *coded atlas* (3.33) with nal_layer_id equal to 0 is a *BLA coded atlas* (3.25)

3.25
broken link access coded atlas frame
BLA coded atlas
BLA atlas frame
IRAP coded atlas (3.75) frame for which each *ACL NAL unit* (3.5) has nal_unit_type equal to NAL_BLA_W_LP, NAL_GBLA_W_LP, NAL_BLA_W_RADL NAL_GBLA_W_RADL, NAL_BLA_N_LP or NAL_GBLA_N_LP.

Note 1 to entry: A BLA coded atlas does not use inter prediction in its decoding process, and could be the first coded atlas in the bitstream in decoding order, or could appear later in the bitstream. Each BLA coded atlas begins a new CAS, and has the same effect on the decoding process as an instantaneous decoding refresh (IDR) coded atlas. However, a BLA coded atlas contains syntax elements that specify a non-empty DAB. When a BLA coded atlas frame for which each ACL NAL unit has nal_unit_type equal to NAL_BLA_W_LP or NAL_GBLA_W_LP, it may have associated random access skipped leading (RASL) coded atlas frames, which are not output by the decoder and may not be decodable, as they may contain references to atlas frames that are not present in the bitstream. When a BLA coded atlas frame for which each ACL NAL unit has nal_unit_type equal to NAL_BLA_W_LP or NAL_GBLA_W_LP, it may also have associated RADL coded atlas frames, which are specified to be decoded. When a BLA coded atlas frame for which each ACL NAL unit has nal_unit_type equal to NAL_BLA_W_RADL or NAL_GBLA_W_RADL, it does not have associated RASL coded atlas frames but may have associated random access decodable leading (RADL) coded atlas frames. When a BLA coded atlas frame for which each ACL NAL unit has nal_unit_type equal to NAL_BLA_N_LP or NAL_GBLA_N_LP, it does not have any associated leading coded atlas frames.

3.26
byte-aligned
positioned as an integer multiple of 8 bits from the position of the first bit in the *bitstream* (3.22)

3.27
byte-aligned position
position in a *bitstream* (3.22) that is *byte-aligned* (3.26)

3.28
byte-aligned byte
byte (3.23) that appears in a position in a *bitstream* (3.22) that is *byte-aligned* (3.26)

3.29
Cartesian coordinates
three scalars (x, y, z) with finite precision and dynamic range that indicate the location of a point relative to a fixed reference point (the origin)

3.30
clean random access access unit
CRA access unit
access unit in which the *coded atlas* (3.33) with nal_layer_id equal to 0 is a *CRA coded atlas* (3.31)

3.31
clean random access coded atlas frame
CRA coded atlas
CRA atlas frame
IRAP coded atlas (3.75) for which each *ACL NAL unit* (3.5) has nal_unit_type equal to NAL_CRA or NAL_GCRA

Note 1 to entry: A CRA coded atlas does not use inter prediction in its decoding process, and could be the first coded atlas in the bitstream in decoding order, or could appear later in the bitstream. A CRA coded atlas could have associated RADL or RASL coded atlas frames. When a CRA coded atlas has NoOutputBeforeRecoveryFlag equal to 1, the associated RASL coded atlas frames are not output by the decoder, because they might not be decodable, as they could contain references to coded atlas frames that are not present in the bitstream.

3.32
codec
specification, device, or system that specifies or uses well defined instructions for encoding or decoding a digital data, i.e. image or video, stream or signal

3.33
coded atlas
coded atlas frame
coded representation of an *atlas* (3.3)

3.34
coded atlas access unit
set of atlas *NAL units* (3.79) that are associated with each other according to a specified classification rule, are consecutive in decoding order, and contain all atlas *NAL units* (3.79) pertaining to one particular output time

3.35
coded atlas sequence
CAS
sequence of *coded atlas access units* (3.34), in decoding order, of an *IRAP coded atlas access unit* (3.74) with NoOutputBeforeRecoveryFlag equal to 1, followed by zero or more *coded atlas access units* (3.34) that are not *IRAP coded atlas access units* (3.74) with NoOutputBeforeRecoveryFlag equal to 1, including all subsequent *access units* (3.34) up to but not including any subsequent *coded atlas access unit* (3.34) that is an *IRAP coded atlas access unit* (3.74) with NoOutputBeforeRecoveryFlag equal to 1.

Note 1 to entry: An IRAP coded atlas access unit may be an IDR coded atlas access unit, a BLA coded atlas access unit, or a CRA coded atlas access unit. The value of NoOutputBeforeRecoveryFlag is equal to 1 for each IDR coded atlas access unit, each BLA coded atlas access unit, and each CRA coded atlas access unit that is the first coded atlas access unit in the atlas bitstream in decoding order, is the first coded atlas access unit that follows an end of sequence NAL unit in decoding order, or has HandleCraAsBlaFlag equal to 1.

3.36
coded common atlas access unit
set of common atlas *non-ACL NAL units* (3.81) that are associated with each other according to a specified classification rule, are consecutive in decoding order, and contain all common atlas *NAL units* (3.79) pertaining to one particular output time

3.37
coded common atlas frame
coded representation of a *common atlas frame* (3.44)

3.38
coded common atlas sequence
CCAS
sequence of *coded common atlas access units* (3.36), in decoding order, of an *IRAP coded common atlas access unit* (3.77), followed by zero or more *coded common atlas access units* (3.36) that are not *IRAP coded common atlas access units* (3.77), including all subsequent *access units* (3.36) up to but not including any subsequent *coded common atlas access unit* (3.36) that is an *IRAP coded common atlas access unit* (3.77)

3.39
coded volumetric frame
collection of coded representations of an *atlas* (3.3), *occupancy* (3.86), *geometry access unit* (3.62), and, for each available *attribute* (3.14), *attribute access unit* (3.15), pertaining to one particular time instance

3.40
coded V3C sequence
CVS
sequence of *V3C sub-bitstreams* (3.134) identified and separated by appropriate delimiters, required to start with a *VPS* (3.135), included in at least one *V3C unit* (3.136) or provided through external means, and contains one or more V3C units that can be factored into *V3C composition units* (3.132), where the first V3C composition unit is a *V3C IRAP composition unit* (3.133)

3.41
coded representation
data element as represented in its coded form

3.42
coded sub-bitstream sequence
sub-bitstream IRAP composition unit (3.111) followed by zero or more *sub-bitstream composition units* (3.110)

3.43
coded V3C component
coded representation of a *V3C component* (3.130)

3.44
common atlas frame
CAF
common information (e.g. projection parameters) that applies to all *atlas frames* (3.7) in a *volumetric frame* (3.139)

3.45
common atlas sequence
collection of *common atlas frames* (3.44)

3.46
common atlas sub-bitstream
extracted *sub-bitstream* (3.109) from the *V3C bitstream* (3.129) containing a part of a common atlas NAL bitstream

3.47
component bitstream
bitstream (3.22) representing a *V3C component* (3.130)

3.48
component sub-bitstream
portion of *component bitstream* (3.44)

3.49
composition time
time or time period at which a frame needs to be composed, used for reconstruction, or presented

3.50
composition time index
index to an ordered list of *composition times* (3.49)

3.51
composition unit
partition of a *bitstream* (3.22) that has a certain *composition time* (3.49)

3.52
decoder under test
DUT
decoder that is tested for conformance to this document by operating the hypothetical stream scheduler to deliver a conforming *bitstream* (3.22) to the decoder and to the hypothetical reference decoder and comparing the values and timing or order of the output of the two decoders

3.53
decoding unit
sub-set of a *coded atlas access unit* (3.34) consisting of one or more ACL NAL units in a *coded atlas access unit* (3.34) and the *associated non-ACL NAL units* (3.2)

3.54
enhanced occupancy mode
EOM
patch coding mode where a *patch* (3.88) is associated with enhanced occupancy information

3.55
EOM coded points
coded representation of 3D points located at intermediate depth positions for which *geometry* (3.61) values are stored as codewords in the *occupancy frame* (3.87) and their corresponding attributes values are stored in additional patches, referred to as *EOM patches* (3.56), in the *attribute frames* (3.16)

3.56
EOM patch
patch (3.88) with *patch mode* (3.90) equal to I_EOM, P_EOM, or P_SKIP

3.57
EOM patch type
patch type (3.91) indicating an *EOM patch* (3.56)

3.58
essential supplemental enhancement information
ESEI
SEI (3.118) that is deemed as essential by the decoding process and should not be ignored or discarded

3.59
essential supplemental enhancement information NAL unit
ESEI NAL unit
NAL unit (3.79) corresponding to an *ESEI* (3.58) and has nal_unit_type equal to NAL_PREFIX_ESEI or NAL_SUFFIX_ESEI

3.60
flag
variable or single-bit syntax element that can take one of the two possible values: 0 and 1

3.61
geometry
set of *Cartesian coordinates* (3.29) associated with a *volumetric frame* (3.139)

3.62
geometry access unit
collection of *geometry maps* (3.64) and auxiliary geometry frames, if present, corresponding to the same time instance

3.63
geometry frame
2D array created through the aggregation of the *geometry* (3.61) information associated with each *patch* (3.88)

3.64
geometry map
geometry frame (3.63) containing geometry *patch* (3.88) information projected at a particular depth

3.65
global broken link access coded atlas frame
GBLA atlas frame
IRAP coded atlas (3.75) frame for which each *ACL NAL unit* (3.5) has nal_unit_type equal to NAL_GBLA_W_LP, NAL_GBLA_W_RADL, or NAL_GBLA_N_LP respectively

3.66
global clean random access coded atlas frame
GCRA atlas frame
IRAP coded atlas (3.75) frame for which each *ACL NAL unit* (3.5) has nal_unit_type equal to NAL_GCRA

3.67
global instantaneous decoding refresh coded atlas frame
GIDR atlas frame
coded atlas (3.33) for which each *ACL NAL unit* (3.5) has nal_unit_type equal to NAL_GIDR_W_RADL, NAL_GBLA_N_LP, or NAL_GCRA, or in the range of NAL_GBLA_W_LP to NAL_GBLA_N_LP, inclusive, and specify a random access association between the current *coded atlas* and its corresponding coded video frames at the same composition time

3.68
hypothetical reference decoder
HRD
hypothetical decoder model that specifies constraints on the variability of conforming atlas *NAL unit* (3.79) streams or conforming *coded atlas* (3.33) sample streams that an encoding process may produce

3.69
hypothetical stream scheduler
HSS
hypothetical delivery mechanism used for checking the conformance of an *atlas sub-bitstream* (3.12) or a decoder with regards to the timing and data flow of the input of an *atlas sub-bitstream* (3.12) into the *hypothetical reference decoder* (3.68)

3.70
instantaneous decoding refresh coded atlas access unit
IDR coded atlas access unit
access unit in which the *coded atlas* (3.33) with nal_layer_id equal to 0 is an *IDR coded atlas* (3.71)