INTERNATIONAL STANDARD

ISO/IEC 23090-2

Third edition
2023-06

Information technology — Coded representation of immersive media —

Part 2: Omnidirectional media format

Technologies de l'information — Représentation codée de média immersifs —
Partie 2: Format de média omnidirectionnel

Reference number
ISO/IEC 23090-2:2023(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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC 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) or the IEC list of patent declarations received (see patents.iec.ch).

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 third edition cancels and replaces the second edition (ISO/IEC 23090-2:2021), which has been technically revised.

The main changes are as follows:

— The following profiles for the Versatile Video Coding (VVC) standard (Rec. ITU-T H.266 | ISO/IEC 23090-3) have been added:

  — VVC-based simple tiling OMAF video profile,

  — VVC-based viewport-independent OMAF video profile,

  — OMAF VVC image profile, and

  — CMAF media profile for the VVC-based viewport-independent OMAF video profile.

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

When omnidirectional media content is consumed with a head-mounted display and headphones, only the parts of the media that correspond to the user's viewing orientation are rendered, as if the user were in the spot where and when the media was captured. One of the forms of omnidirectional media applications is omnidirectional video, also known as 360° video. Omnidirectional video is typically captured by multiple cameras that cover the entire sphere or at least a large part of the sphere. Compared to traditional media application formats, the end-to-end technology for omnidirectional video (from capture to playback) is more easily fragmented due to various capturing and video projection technologies. From the capture side, there exist many different types of cameras capable of capturing 360° video, and on the playback side there are many different devices that are able to playback 360° video with different processing capabilities. To avoid fragmentation of omnidirectional media content and devices, a standardized format for omnidirectional media applications is specified in this document.

This document defines a media format that enables omnidirectional media applications, focusing on 360° video, images, and audio, as well as associated timed text. What is specified in this document includes (but is not limited to):

— a coordinate system that consists of a unit sphere and three coordinate axes, namely the X (back-to-front) axis, the Y (lateral, side-to-side) axis, and the Z (vertical, up) axis;

— projection and rectangular region-wise packing methods that may be used for conversion of a spherical video sequence or image into a two-dimensional rectangular video sequence or image, respectively;

— storage of omnidirectional media and the associated metadata using the ISO Base Media File Format (ISOBMFF) as specified in ISO/IEC 14496-12;

— storage of video or image overlays and the associated metadata using ISOBMFF;

— encapsulation, signalling, and streaming of omnidirectional media and overlays in a media streaming system, e.g. dynamic adaptive streaming over HTTP (DASH) as specified in ISO/IEC 23009-1 or MPEG media transport (MMT) as specified in ISO/IEC 23008-1;

— media profiles and presentation profiles that provide conformance points for media codecs as well as media coding and encapsulation configurations that may be used for compression, streaming, and playback of the omnidirectional media content;

— toolset brands that provide conformance points for functionalities beyond plain 360° video, images and audio.

The International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that compliance with this document may involve the use of a patent.

ISO and IEC take no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has assured ISO and IEC that he/she is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with ISO and IEC. Information may be obtained from the patent database available at www.iso.org/patents.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those in the patent database. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Information technology — Coded representation of immersive media —

Part 2: Omnidirectional media format

1 Scope

This document specifies the omnidirectional media format for coding, storage, delivery and rendering of omnidirectional media, including video, images, audio and timed text. Omnidirectional image or video can contain graphics elements generated by computer graphics but encoded as image or video. Multiple viewpoints, each corresponding to an omnidirectional camera, are supported. The document also specifies storage and delivery of overlay images or video intended to be rendered over the omnidirectional background image or video.

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.

ISO/IEC 10918-1, *Information technology — Digital compression and coding of continuous-tone still images — Part 1: Requirements and guidelines*

ISO/IEC 14496-1:2010, *Information technology — Coding of audio-visual objects — Part 1: Systems*

ISO/IEC 14496-3:2019, *Information technology — Coding of audio-visual objects — Part 3: Advanced audio coding*

Rec. ITU-T H.264 (08/21) | ISO/IEC 14496-10:2020, *Information technology — Coding of audio-visual objects — Part 10: Advanced video coding*

ISO/IEC 14496-12:2022, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 14496-14, *Information technology — Coding of audio-visual objects — Part 14, MP4 file format*

ISO/IEC 14496-15:2022, *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 14496-30:2018, *Information technology — Coding of audio-visual objects — Part 30: Timed text and other visual overlays in ISO base media file format*

ISO/IEC 23000-19:—[1], *Information technology — Multimedia application format (MPEG-A) — Part 19: Common media application format (CMAF) for segmented media*

ISO/IEC 23000-22:2019, *Information technology — Multimedia application format —Part 22 Multi-image application format (MIAF)*

Rec. ITU-T H.274 | ISO/IEC 23002-7, *Information technology — MPEG video technologies — Part 7: Versatile supplemental enhancement information messages for coded video bitstreams*

ISO/IEC 23003-3:2020, *Information technology — MPEG audio technologies — Part 3: Unified speech and audio coding*

ISO/IEC 23003-4:2020, *Information technology — MPEG audio technologies — Part 4: Dynamic range control*

ISO/IEC 23008-1:2017, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 1: MPEG media transport (MMT)*

Rec. ITU-T H.265 (11/19) | ISO/IEC 23008-2:2020, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

ISO/IEC 23008-3:2022, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio*

ISO/IEC 23008-12, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 12: Image file format*

ISO/IEC 23009-1:2022, *Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats*

Rec. ITU-T H.266 (08/20) | ISO/IEC 23090-3:2021, *Information technology — Coded representation of immersive media — Part 3: Versatile video coding*

ISO/IEC 23091-2, *Information technology — Coding-independent code points — Part 2: Video*

ISO/IEC 23091-3, *Information technology — Coding-independent code points — Part 3: Audio*

IETF BCP 47, *Tags for Identifying Languages*

IETF Internet Standard 66, *Uniform Resource Identifier (URI): Generic Syntax*

IETF RFC 6381, *MIME Codecs and Profiles*

W3C Candidate Recommendation, *WebVTT: The Web Video Text Tracks Format*

W3C Recommendation, *TTML Profiles for Internet Media Subtitles and Captions 1.0.1 (IMSC1)*

W3C Recommendation, *XML schema part 1: Structures*

W3C Recommendation, *XML schema part 2: Datatypes*

[1] Under preparation. Stage at the time of publication: ISO/IEC DIS 23000-19:2022.

W3C Recommendation, *XML Path Language (XPath) 2.0 (Second Edition)*

3 Terms, definitions, abbreviated terms and symbols

3.1 Terms and definitions

For the purposes of this document, the terms and definitions in ISO/IEC 14496-12, ISO/IEC 23008-12 and the following 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/

3.1.1
azimuth

first of the two sphere coordinates describing the location of a point on the sphere

Note 1 to entry: Azimuth and elevation are specified in subclause 5.1.

3.1.2
azimuth circle

circle on the sphere connecting all points with the same azimuth value

Note 1 to entry: An *azimuth circle* is always a *great circle* (3.1.22).

3.1.3
background visual media

piece of *visual media* (3.1.66) on which an *overlay* (3.1.34) is superimposed

3.1.4
circular image

image captured with a *fisheye lens* (3.1.18)

3.1.5
closed scheme type

scheme type (3.1.49) that clearly specifies which transformations are allowed and does not allow future extensions

3.1.6
common reference coordinate system

3D Cartesian coordinate system with the centre being (X, Y, Z) equal to (0, 0, 0), used as the reference coordinate system for all *viewpoints* (3.1.63) within a *viewpoint group* (3.1.64)

3.1.7
composition-aligned sample

sample in a track that is associated with another track, the sample has the same composition time as a particular sample in the another track, or, when a sample with the same composition time is not available in the another track, the closest preceding composition time relative to that of a particular sample in the another track

3.1.8
composition picture

picture that is suitable to be presented and is obtained from the decoding outputs of *composition-aligned samples* (3.1.7) of all tracks of a 2D spatial relationship track group by arranging them spatially as specified by the semantics of the 2D spatial relationship track group

3.1.9
constituent picture

such part of a spatially frame-packed stereoscopic picture that corresponds to one view, or a picture itself when frame packing is not in use or the temporal interleaving frame packing arrangement is in use

3.1.10
content coverage

one or more *sphere regions* (3.1.51) that are covered by the content represented by the track, an image item, or a *composition picture* (3.1.8)

3.1.11
elevation

second of the two sphere coordinates describing the location of a point on the sphere

Note 1 to entry: Azimuth and elevation are specified in subclause 5.1.

3.1.12
elevation circle

circle on the sphere connecting all points with the same elevation value

Note 1 to entry: When the elevation is zero, an *elevation circle* is also a *great circle* (3.1.22). This coincides with the equator on Earth.

3.1.13
encoded tile sequence

coded representation of a *tile sequence* (3.1.56) that has the capability to be merged with other encoded tile sequences in coded domain without decoding mismatch by rewriting only header data

3.1.14
extractor track

track that has untransformed sample entry type equal to `'hvc2'`, `'avc2'`, or `'avc4'` and contains one or more `'scal'` track references

3.1.15
field of view

extent of the observable world in captured/recorded content or in a physical display device

3.1.16
file decoder

collective term for file/segment decapsulation and decoding of video, audio or image bitstreams

3.1.17
file decoding process

process specified as a part of a media profile specification that takes as input a set of ISOBMFF tracks or items and derives either a decoded pictures or audio samples, and rendering metadata for them; or a fully rendered audio scene in the reference system

3.1.18
fisheye lens

wide-angle camera lens that usually captures an approximately hemispherical *field of view* (3.1.15) and projects it as a *circular image* (3.1.4)

3.1.19
fisheye omnidirectional video
fisheye omnidirectional image

omnidirectional media (3.1.32) where *circular images* (3.1.4) of *fisheye lenses* (3.1.18) are spatially arranged onto picture(s)

3.1.20
fisheye video

video captured by *fisheye lenses* (3.1.18)

3.1.21
global coordinate axes

coordinate axes that are associated with audio, video, and images representing the same acquisition position and intended to be rendered together

Note 1 to entry: Coordinate axes are specified in subclause 5.1.

Note 2 to entry: The origin of the global coordinate axes is usually the same as the centre point of a device or rig used for omnidirectional audio/video acquisition as well as the position of the observer's head in the three-dimensional space in which the audio and video tracks are located.

Note 3 to entry: In the absence of the initial viewing orientation metadata (see subclause 7.7.4 for tracks or subclause 7.9.9 for image items), the initial viewing orientation should be inferred to be equal to (0, 0, 0) for (`centre_azimuth`, `centre_elevation`, `centre_tilt`) relative to the global coordinate axes.

3.1.22
great circle

intersection of the sphere and a plane that passes through the centre point of the sphere

Note 1 to entry: A *great circle* is also known as an orthodrome or Riemannian circle.

Note 2 to entry: The centre of the sphere and the centre of a *great circle* are co-located.

3.1.23
guard band

area that is not rendered but may be used to improve the rendering quality to avoid or mitigate visual artifacts such as seams

Note 1 to entry: Guard bands in *packed pictures* (3.1.37) are associated with *packed regions* (3.1.38) as described in subclause 7.5.3.

3.1.24
header data

high-level syntax structures that do not contain arithmetic entropy-coded syntax elements

Note 1 to entry: In HEVC, parameter sets and slice segment headers can be regarded as header data. In VVC bitstreams with one or more independent VVC subpictures, parameter sets can be regarded as header data.

3.1.25
independent VVC subpicture

subpicture, as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3, that has sps_subpic_treated_as_pic_flag[i] equal to 1 and sps_loop_filter_across_subpic_enabled_pic_flag[i] equal to 0

Note 1 to entry: An independent VVC subpicture is treated like a picture in the VVC decoding process and loop filtering across the boundaries of an independent VVC subpicture is disabled.

3.1.26
local coordinate axes

coordinate axes obtained after applying rotation to the *global coordinate axes* (3.1.21)

3.1.27
mesh omnidirectional video

omnidirectional media (3.1.32) where rectangular regions of two-dimensional pictures are mapped to mesh elements of a three-dimensional mesh

3.1.28
OMAF base track

track whose samples are resolved by merging samples of the referenced *OMAF tile tracks* (3.1.30)

Note 1 to entry: An HEVC tile base track as specified in ISO/IEC 14496-15, an *extractor track* (3.1.14), and a VVC merge base track as specified in ISO/IEC 14496-15 that reference OMAF tile tracks are examples of OMAF base tracks.

3.1.29
OMAF player

application responsible for receiving files or segments or accessing files locally, decapsulating these files and segments, decoding the audio, video, image or timed text bitstreams, rendering the audio, video, image or timed text and applying a strategy for receiving the files or segments based on the *viewport* (3.1.65) information

3.1.30
OMAF tile bitstream

coded video data that represents a rectangular subset of the original video content and has the capability to be merged with other OMAF tile bitstreams in coded domain without decoding mismatch by rewriting only *header data* (3.1.24), where a decoding mismatch refers to the value of any pixel having a different value when decoded prior to merging or subsequently to merging

3.1.31
OMAF tile track

track that represents a rectangular subset of the original video content and has the capability to be merged with other OMAF tile tracks in coded domain without decoding mismatch by rewriting only header data, where a decoding mismatch refers to the value of any pixel having a different value when decoded prior to merging or subsequently to merging

Note 1 to entry: An HEVC tile track as specified in ISO/IEC 14496-15 and a VVC subpicture track as specified in ISO/IEC 14496-15 are examples of an OMAF tile track.

3.1.32
omnidirectional media

media such as image or video and its associated audio that enable rendering according to the user's *viewing orientation* (3.1.60), if consumed with a head-mounted device, or according to user's desired *viewport* (3.1.65), otherwise, as if the user was in the spot where and when the media was captured

3.1.33
open-ended scheme type

scheme type (3.1.49) that allows future extensions

3.1.34
overlay

piece of *visual media* (3.1.66) rendered over omnidirectional video or image item or over a *viewport* (3.1.65)

3.1.35
overlay source

piece of *visual media* (3.1.66) used as the content for an *overlay* (3.1.34)

3.1.36
overlay source region

overlay source (3.1.35) specified as a rectangular region within decoded pictures

3.1.37
packed picture

picture that is represented as a coded picture in the coded video bitstream

Note 1 to entry: A *packed picture* may result from *region-wise packing* (3.1.46) of a *projected picture* (3.1.40).

3.1.38
packed region

region in a *packed picture* (3.1.37) that is mapped to a *projected region* (3.1.41) as specified by the region-wise packing signalling

3.1.39
projected omnidirectional video
projected omnidirectional image

omnidirectional media (3.1.32) where the relation of two-dimensional picture coordinates to *sphere coordinates* (3.1.50) is specified using mathematical projection equation(s)