DRAFT INTERNATIONAL STANDARD

ISO/IEC DIS 23000-20

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

Voting begins on:
2017-07-13

Voting terminates on:
2017-10-04

Information technology — Multimedia application format (MPEG-A) —

Part 20:
Omnidirectional media application format

Technologies de l'information — Format pour application multimédia (MPEG-A)

ICS: 35.040.40

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENT AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

This document is circulated as received from the committee secretariat.



Reference number
ISO/IEC DIS 23000-20:2017(E)

© ISO/IEC 2017



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2017, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

Contents

Page

1 Scope ... 8
2 Normative references ... 8
3 Terms, definitions, abbreviated terms, and conventions ... 9
3.1 Terms and definitions ... 9
3.2 Abbreviations and acronyms ... 10
3.3 Arithmetic operators and mathematical functions ... 10
4 OMAF architecture ... 11
4.1 Overall architecture ... 11
4.2 Conformance and interoperability ... 14
5 Omnidirectional video projection and region-wise packing ... 14
5.1 Projection structure and coordinate system ... 14
5.2 Omnidirectional projection formats ... 15
5.3 Conversion between spherical coordinate systems of different orientations ... 16
5.4 Conversion of sample locations for rectangular region-wise packing ... 16
6 Omnidirectional fisheye video ... 17
6.1 Introduction ... 17
6.2 Syntax ... 17
6.3 Semantics ... 18
7 Omnidirectional media storage and metadata signaling in the ISOBMFF ... 23
7.1 Generic extensions to the ISOBMFF ... 23
7.2 Restricted video schemes for omnidirectional video ... 25
7.3 Sub-picture composition track grouping applied to omnidirectional projected video ... 34
7.4 Timed metadata for regions on sphere ... 34
7.5 Storage of images ... 39
7.6 Signalling of region-wise quality ranking ... 41
8 Omnidirectional media encapsulation and signaling in DASH ... 43
8.1 Architecture of DASH delivery in OMAF ... 43
8.2 Usage of DASH in OMAF ... 44
9 Omnidirectional media encapsulation and signaling in MMT ... 46
9.1 Architecture of MMT delivery in OMAF ... 46
9.2 OMAF Signaling in MPEG Composition Information ... 48
10 Media profiles ... 48

© ISO/IEC 2017 – All rights reserved

11 Presentation profiles 48
Annex A. Viewport dependent omnidirectional video processing (informative) 49
A.1. General 49
A.2. Omnidirectional video streaming based on viewport dependent video coding 49
Annex B. DASH MPD examples (informative) 58
B.1. Introduction 58
B.2. Single stream delivery 58
B.3. Sub-picture bitstream delivery 59

© ISO/IEC 2017 – All rights reserved

Editors' notes on DIS of Omnidirectional Media Format.

List of adoptions integrated:

- m40533: It was agreed to add the support of linear interpolation functionality for sphere domain ROI data on sample level. Agreed the syntax and semantics, in principle, in part 2.8 of the last version of m40501, and include the definition of sync sample of the timed metadata track being a sample for which the corresponding media sample is a sync sample, and further constrain that the value of the interpolation field shall indication no interpolation for a sync sample.
- m40603: It was agreed that the overlapping sub-picture scheme described in the document is interesting, and was agreed to add a description of the coding scheme and file format signalling mechanism (to the extent similarly as for the current non-overlapping sub-picture based scheme) to Annex A.
- m40563
    - (AAC) It was agreed to adopt the text in v2 of this document with the understanding of 4 channels.
    - (AVC) It was agreed to add the support of AVC High profile, Level 5.1, and define separate media profiles for HEVC and AVC.
- Sub-picture tracks and related
    - Signalling of full picture size: Use the existing track group design in ISOBMFF, need to define a new track grouping type, and signal the full picture size in each and every sub-picture track, in luma samples. Agreed.
    - Signalling of position and size of each sub-picture in 2D domain: Use the existing track group design in ISOBMFF, need to define a new track grouping type, in luma samples. Agreed.
    - The relationship among the sub-picture tracks corresponding to the "full picture bitstream": Use the existing track group design in ISOBMFF: need to define a new grouping type. Agreed.
    - Resolving the "issue" of media track referencing by timed metadata tracks: Refer to all the sub-picture tracks using a new reference type that applies to the content represented by all the referred tracks, and clarify that 'cdsc' applies to the tracks individually when multiple media tracks are referred to. Agreed.
    - Sub-picture track coverage signalling of only one sphere region: Signalling of coverage info of the video piece carried by a sub-picture track, using the CoverageInformationBox() defined and contained as currently specified in the OMAF CD. Agreed to clarify the semantics of CoverageInformationBox() to apply only to the video content represented by the track only, unless it is contained in a tile base track.
    - A specific box in the 'resv' sample entry to indicate that a sub-picture track is not sufficient to be presented alone: Agreed to signal an indication, in the visual sample entry, of a sub-picture track, that the sub-picture track should not be presented alone, if so. Use a box that contains no syntax element for this purpose.
- m40501: 2.6: The second point was delegated to the editors.
- m40454: Remove the constraint on the sum of center_pitch + ver_range/2. Agreed.
- m40713: It was agreed to use a different restricted scheme type for fisheye. This way, there is no need to specify a DASH signalling for fisheye indication.
- m40728: Sections 2.1, 2.2, 2.3, and 2.4 were agreed to be included into the OMAF DIS text.
- m40523: Section 4.3 was generally agreed, but move the DASH and CMAF aspects to an Annex.

© ISO/IEC 2017 – All rights reserved

- m40730: It was further agreed to add an editor's note to the place of text related to profiles as noted under m40523 in the DIS to refer to the Profile under Consideration and the output document on Submission Template for Profiles in OMAF (ISO/IEC 23090-2).
- m40501: It was agreed that a point on the spherical surface would have yaw and pitch ranges both equal to 0, although the yaw and pitch ranges for a pixel on a picture would not be 0.
- m40501: Change num_regions in RegionOnSphereConfigBox from 16 bits to 8 bits. Agreed.
- m40501: Proposes to constrain that the values of proj_frame_width and proj_frame_height shall not be equal to 0. Agreed.
- m40501: For the syntax in 5.2.1.2 and 7.1.4.2, change the length of the 32-bits fields to 16 bits. Agreed.
- m40297: Decision (Chaired by Madhukar Budagavi): Remove clause 7.2.
- m40557: Sections 9.1 and 9.2 were agreed to be included into the OMAF DIS text, with the last sentence in section 9.2 to be included into an editor's note.
- m40407: All proposals have been agreed except removal of coverage information.
- m40714: Agreed.
- m40408: Agreed.
- m40707: Agreed with the above answers to the open issues.
- m40370: Agreed to adopt the first part.
- m40415: It was agreed to add a note to mention this would be considered. It was decided on other context that all the angle values would be represented using the 16.16 fixed point format. Agreed to change to be unsigned 8.24 fixed point. Define the range of camera X, Y and Z offsets [-1.0~1.0] Agreed. . Editorial updates of semantics of displayed_fov and overlapped_fov. Delegated to editors. Editorial updates of semantics of lens shading compensation parameters. Delegated to editors.
- m40501: Proposing to use 3 bits for transform_type[i] and keep 5 bits reserved. Agreed with the 5 reserved bits go first. Also the order of the 8 types is numbered 0 to 7. Also remove the note.
- m40729: Agreed. This covers the following two items as well:
    - m40231: It was agreed to recommend the use of the 16.16 fixed-point data format to JCT-VC. (This was later confirmed in the joint meeting with JCT-VC and requirements: Angular syntax values should use 32 bits in increments of 2^{-16}.)
    - m40501: Change roll from unsinged to signed. Agreed. (probably overridden by the m40231 and m40729 decisions)
- Alignment with JCTVC-AA1005 on projection, including
    - Rremoval of geometry type
    - Removal of equations and tables for chroma handling
- Agreed items in the DoC report in N16823

© ISO/IEC 2017 – All rights reserved

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

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.

© ISO/IEC 2017 – All rights reserved

Introduction

When omnidirectional media content is consumed with a head-mounted display and headphones, only the area of the spherical video and the part of the associated audio that correspond to the orientation of the user's viewing orientation are rendered, as if the user were in the spot where and when the video and audio were captured. One of the most popular forms of omnidirectional media applications is omnidirectional video, also known as 360° video. Omnidirectional video is typically captured by multiple cameras that cover up to 360° of the scene. 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 can 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 part of ISO/IEC 23090.

This specification defines a media application format that enables omnidirectional media applications, focusing on 360° video and associated audio. It firstly specifies a list of projection methods that can be used for conversion of a spherical video into a two-dimensional rectangular video, followed by how to store omnidirectional media and the associated metadata using the ISO base media file format (ISOBMFF) and how to encapsulate, signal, and stream omnidirectional media in a media streaming system, e.g., dynamic adaptive streaming over HTTP (DASH), and finally which video and audio codecs as well as media coding configurations can be used for compression and playback of the omnidirectional media signal.

© ISO/IEC 2017 – All rights reserved

Information technology — Multimedia application format (MPEG-A) -- Part 20: Omnidirectional media application format

1 Scope

This part of ISO/IEC 23000 specifies the omnidirectional media format for coding, storage, delivery, and rendering of omnidirectional images and video and the associated audio.

Omnidirectional video may contain graphics elements generated by computer graphics. At the receiver side, it is assumed that the user sits in the center of the sphere and looks towards outside.

NOTE: In this version of this Specification, only 3 Degrees of Freedom (3DOF) is supported. In other words, purely translational movement of the user does not result in different omnidirectional media being rendered to the user. For 3DOF support with stereoscopic video, when the user rolls his/her head, there can be a stereoscopic rendering issue.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

These normative references are intended to include corrigenda and amendments available at the time of use.

[3DA]	ISO/IEC 23008-3, Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D Audio, Second Edition
[AVC]	ISO/IEC 14496-10, Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding
[CICP]	ISO/IEC 23001-8:2016 -- MPEG systems technologies -- Part 8: Coding-independent code points
[CMAF]	ISO/IEC 23000-19, Information technology — Part 19: Common media application format (CMAF)
[DASH]	ISO/IEC 23009-1, Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats
[DASH DAM 4]	ISO/IEC 23009-1:2014 DAM 4 Segment Independent SAP Signalling (SISSI), MPD chaining, MPD reset and other extensions
[HEVC]	ISO/IEC 23008-2, Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding
[ISOM]	ISO/IEC 14496-12:2015 Information technology – Coding of audio-visual objects – Part 12: ISO Base Media File Format
[MMT]	ISO/IEC 23008-1, Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 1: MPEG media transport (MMT)

© ISO/IEC 2017 – All rights reserved

3 Terms, definitions, abbreviated terms, and conventions

3.1 Terms and definitions

field of view the extent of the observable world in captured/recorded content or in a physical display device (e.g., a Head Mounted Display), which can be either horizontal (Horizontal Field of View) or vertical (Vertical Field of View).

global coordinate axes coordinate axes, as defined in clause 5.1, that are associated with audio, video, and images representing the same acquisition position and intended to be rendered together

NOTE 1: The origin of the global coordinate axes is usually the same as the center 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 2: In the absence of the initial viewpoint region-on-sphere metadata (see 7.4.4), the playback should be started using the orientation (0, 0, 0) in (yaw, pitch, roll) relative to the global coordinate axes.

great circle intersection of the sphere and a plane that passes through the center point of the sphere

NOTE 1: A *great circle* is also known as an orthodrome or Riemannian circle.

NOTE 2: The center of the sphere and the center of a *great circle* are co-located.

guard band area in a packed picture that is not rendered but may be used to improve the rendered part of the packed picture to avoid or mitigate visual artifacts such as seams

NOTE: Guard bands are associated with regions as described in 7.2.5.

horizontal range horizontal *field of view* through the center point of the region, if a region is specified by four great circles, or otherwise a range of yaw angles

local coordinate axes coordinate axes, as defined in clause 5.1, that are associated with the *projection structure* of a video track or an image item and indicate the orientation of the *projection structure* relative to the *global coordinate axes*

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

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

NOTE: A *packed picture* may result from *region-wise packing* of a *projected picture*.

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

NOTE: A *pitch circle* is not necessarily a *great circle* like a latitude circle on the earth.

projected picture picture that has a representation format specified by an omnidirectional video projection format specified in clause 5.2

projection inverse of the process by which the samples of a *projected picture* are mapped to a set of positions identified by a set of coordinates on a *projection structure*

© ISO/IEC 2017 – All rights reserved

projection structure three-dimensional structure consisting of one or more surfaces on which the samples of a *projected picture* can be mapped to a set of positions identified by a set of coordinates

region-wise packing region-wise transformation, resizing, and relocating of a *projected picture* when generating the *packed picture* before encoding

rendering process of generating a displayed audio-visual content from the decoded audio-visual data according to the user's viewing orientation, if consumed with a head-mounted device, or according to user's desired viewport, otherwise

time-parallel sample sample in an associated track that has the same or, when a sample with the same decoding time is not available, the closest preceding decoding time relative to that of a particular sample in a track

vertical range vertical *field of view* through the center point of the region, if a region specified by four *great circles*, or otherwise a range of pitch angles

viewport part of the spherical video that is currently displayed and viewed by the user

NOTE: The size of the *viewport* can be smaller than or equal to the *field of view*.

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

NOTE: A *yaw circle* is always a *great circle* like a longitude line on the earth.

3.2 Abbreviations and acronyms

2D	two-dimensional
CICP	Coding Independent Code-Points (ISO/IEC 23001-8)
CMAF	Common Media Application Format (ISO/IEC 23000-19)
DASH	MPEG Dynamic Adaptive Streaming over HTTP (ISO/IEC 23009-1)
FOV	Field Of View
ISOBMFF	ISO base media file format (ISO/IEC 14496-12)
HEVC	High Efficiency Video Coding (ISO/IEC 23008-2)
MCTS	motion-constrained tile set
MMT	MPEG media transport (ISO/IEC 23008-1)
OMAF	Omnidirectional Media Format (ISO/IEC 23090-2)

3.3 Arithmetic operators and mathematical functions

+	Addition
−	Subtraction (as a two-argument operator) or negation (as a unary prefix operator)
*	Multiplication, including matrix multiplication

© ISO/IEC 2017 – All rights reserved