ISO/IEC 23090-2:2019-01 (E)

Information technology - Coded representation of immersive media - Part 2: Omnidirectional media format

Contents

Page

Foreword vii

Introduction viii

1 Scope 1

2 Normative references 1

3 Terms, definitions, abbreviated terms, and conventions 2

3.1 Terms and definitions 2

3.2 Abbreviated terms 6

3.3 Arithmetic operators and mathematical functions 7

3.4 Order of operation precedence 8

3.5 Range notation 9

3.6 Variables 9

3.7 Processes 9

4 Overview 9

4.1 Organization of this document 9

4.2 Overall architecture for omnidirectional media with projected video 10

4.2.1 Overview 10

4.2.2 Stitching, rotation, projection, and region-wise packing 11

4.3 Overall architecture for omnidirectional media with fisheye video 12

4.4 Conformance and interoperability 13

4.4.1 General 13

4.4.2 Media profiles 14

4.4.3 Presentation profiles 15

4.4.4 Summary of referenceable code points 16

5 Omnidirectional video projection and region-wise packing 19

5.1 Coordinate system 19

5.2 Omnidirectional projection formats 20

5.2.1 General 20

5.2.2 Equirectangular projection for one sample location 20

5.2.3 Cubemap projection for one sample location 21

5.3 Conversion from the local coordinate axes to the global coordinate axes 23

5.4 Region-wise packing formats 24

5.4.1 General 24

5.4.2 Conversion of one sample location for rectangular region-wise packing 24

6 Fisheye omnidirectional video 25
6.1 General 25
6.2 FisheyeVideoEssentialInfoStruct syntax structure 26
6.2.1 Syntax 26
6.2.2 Semantics 26
6.3 FisheyeVideoSupplementalInfoStruct syntax structure 29
6.3.1 Syntax 29
6.3.2 Semantics 30
7 Omnidirectional media storage and metadata signalling in the ISOBMFF 33
7.1 Generic extensions to the ISOBMFF 33
7.1.1 Stereoscopic video track grouping 33
7.1.2 Indication of `track_group_id` uniqueness 34
7.1.3 Updated semantics of `track_IDs` of the track reference box 34
7.1.4 Indication of a track not intended to be presented alone 34
7.1.5 Timed metadata tracks 34
7.1.6 Compatible scheme type box 35
7.1.7 Multiple transformations for a single transformed media track 35
7.1.8 The `'codecs'` parameter for a transformed media track 35
7.1.9 Track type box 36
7.1.10 Clarifications on the stereo video box 36
7.2 Generic extensions to ISO/IEC 14496-15 37
7.2.1 Alternative extraction source track grouping 37
7.2.2 Tile base track association with coverage information box and timed metadata data track 37
7.3 OMAF-specific extensions to the ISOBMFF 37
7.3.1 Sync samples in timed metadata tracks 37
7.4 OMAF-specific extensions to ISO/IEC 14496-15 37
7.4.1 Coverage information box in a tile base track 37
7.5 Structures and semantics that are common for video tracks and image items 38
7.5.1 Semantics of sample locations within a decoded picture 38
7.5.2 Projection format structure 41
7.5.3 Region-wise packing structure 41
7.5.4 Rotation structure 48
7.5.5 Content coverage structure 48
7.5.6 Sphere region structure 49
7.6 Restricted video schemes for omnidirectional video 51
7.6.1 Scheme types 51
7.6.2 Projected omnidirectional video box 54
7.6.3 Fisheye omnidirectional video box 55
7.6.4 Region-wise packing box 55
7.6.5 Rotation box 56
7.6.6 Coverage information box 56
7.7 Timed metadata for sphere regions 56
7.7.1 General 56
7.7.2 Sample entry 57
7.7.3 Sample format 58
7.7.4 Initial viewing orientation 58
7.7.5 Recommended viewport 59
7.7.6 Timed text sphere location metadata 60
7.8 Signalling of region-wise quality ranking 61
7.8.1 General 61
7.8.2 Spherical region-wise quality ranking 61
7.8.3 2D region-wise quality ranking 63

7.9 Storage of omnidirectional images 65
7.9.1 General 65
7.9.2 Frame packing item property 65
7.9.3 Projection format item property 65
7.9.4 Essential fisheye image item property 66
7.9.5 Supplemental fisheye image item property 67
7.9.6 Region-wise packing item property 67
7.9.7 Rotation item property 68
7.9.8 Coverage information item property 68
7.9.9 Initial viewing orientation item property 69
7.10 Storage of timed text for omnidirectional video 69
7.10.1 General 69
7.10.2 OMAF timed text configuration box 70
7.10.3 IMSC1 tracks 72
7.10.4 WebVTT tracks 73
8 Omnidirectional media encapsulation and signalling in DASH 73
8.1 Architecture of DASH delivery in OMAF 73
8.2 Usage of DASH in OMAF 74
8.2.1 General 74
8.2.2 Signalling of stereoscopic frame packing 74
8.2.3 Carriage of timed metadata 74
8.3 DASH MPD descriptors for omnidirectional media 75
8.3.1 XML namespace and schema 75
8.3.2 Signalling of projection type information 75
8.3.3 Signalling of region-wise packing type 76
8.3.4 Signalling of content coverage 76
8.3.5 Signalling of spherical region-wise quality ranking 79
8.3.6 Signalling of 2D region-wise quality ranking 84
8.3.7 Signalling of fisheye omnidirectional video 88
9 Omnidirectional media encapsulation and signalling in MMT 89
9.1 Architecture of MMT delivery in OMAF 89
9.2 OMAF signalling in MPEG composition information 90
9.3 VR application-specific MMT signalling 90
9.3.1 General 90
9.3.2 MMT signalling 91
10 Media profiles 103
10.1 Video profiles 103
10.1.1 Overview 103
10.1.2 HEVC-based viewport-independent OMAF video profile 103
10.1.3 HEVC-based viewport-dependent OMAF video profile 106
10.1.4 AVC-based viewport-dependent OMAF video profile 109
10.2 Audio profiles 111
10.2.1 Overview 111
10.2.2 OMAF 3D audio baseline profile 111
10.2.3 OMAF 2D audio legacy profile 114
10.3 Image profiles 118
10.3.1 Overview 118
10.3.2 Common specifications for image profiles 119
10.3.3 OMAF HEVC image profile 120
10.3.4 OMAF legacy image profile 121

10.4 Timed text profiles 122
10.4.1 Overview 122
10.4.2 OMAF IMSC1 timed text profile 123
10.4.3 OMAF WebVTT timed text profile 123
11 Presentation profiles 124
11.1 OMAF viewport-independent baseline presentation profile 124
11.1.1 General (informative) 124
11.1.2 ISO base media file format constraints 124
11.2 OMAF viewport-dependent baseline presentation profile 124
11.2.1 General 124
11.2.2 ISO base media file format constraints 124
Annex A (normative) OMAF DASH schema 125
Annex B (normative) DASH integration of media profiles 128
Annex C (normative) CMAF integration of media profiles 134
Annex D (informative) Viewport-dependent omnidirectional video processing 136
Annex E (informative) DASH MPD examples 154
Annex F (informative) MMT signalling examples 158