INTERNATIONAL STANDARD

ISO/IEC 23090-12

First edition
2023-08

Information technology — Coded representation of immersive media —

Part 12:
MPEG immersive video

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

Partie 12: Vidéo immersive MPEG

Reference number
ISO/IEC 23090-12:2023(E)

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*.

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

This document was developed to support compression of immersive video content, in which a real or virtual 3D scene is captured by multiple real or virtual cameras. The use of this document enables storage and distribution of immersive video content over existing and future networks, for playback with 6 degrees of freedom of view position and orientation.

Information technology — Coded representation of immersive media —

Part 12:
MPEG immersive video

1 Scope

This document specifies the syntax, semantics and decoding processes for MPEG immersive video (MIV), as an extension of ISO/IEC 23090-5. It provides support for playback of a three-dimensional (3D) scene within a limited range of viewing positions and orientations, with 6 Degrees of Freedom (6DoF).

2 Normative reference

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 23090-5[1)], *Information technology — Coded Representation of Immersive Media — Part 5: Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC)*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 23090-5 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
atlas sample
position on the rectangular frame associated with an *atlas*

3.2
coded MIV sequence
coded V3C sequence conforming to the constraints specified in this document

3.3
decoding process
process specified in this document that reads a *bitstream* and derives *patch data* and related information that can be used to render a *viewport* (3.22)

3.4
decoding order
order in which syntax elements are processed by the *decoding process* (3.3)

1) Under preparation. Stage at time of publication: ISO/IEC FDIS 23090-5:2023.

3.5
field of view
FOV
angular region of the observable world in captured/recorded content or in a physical display device

3.6
MIV access unit
V3C composition unit that is a set of all *sub-bitstream access units* (3.13) that share the same *decoding order* (3.4) count

3.7
MIV coded sub-bitstream sequence
sub-bitstream IRAP access unit (3.14) followed by zero or more *sub-bitstream access units* (3.13)

Note 1 to entry: A *MIV coded sub-bitstream sequence* is a *coded sub-bitstream sequence* conforming to the constraints specified in this document.

3.8
MIV IRAP access unit
MIV access unit (3.6) for which all *sub-bitstream access units* (3.13) are *sub-bitstream IRAP access units* (3.14)

Note 1 to entry: A *MIV IRAP access unit* is a *V3C IRAP composition unit* conforming to *the constraints specified in this document.*

3.9
multi-plane image
MPI
set of pairs of texture and transparency *attribute frames*, each associated with an implicit constant geometry frame

3.10
renderer
embodiment of a process to create a *viewport* (3.22) from a *volumetric frame* corresponding to a *viewing orientation* (3.19) and *viewing position* (3.20)

3.11
source
one or more video sequences, each containing *geometry* or an *attribute* such as texture and transparency information before encoding

3.12
source view
source (3.11) video material before encoding that corresponds to the format of a *view* (3.15), which may have been acquired by capture of a 3D scene by a real or virtual camera

3.13
sub-bitstream access unit
partition of a *sub-bitstream* that has a certain *decoding order* (3.4) count

Note 1 to entry: A *sub-bitstream access unit* is a *sub-bitstream composition unit*.

3.14
sub-bitstream IRAP access unit
sub-bitstream access unit (3.13) that forms an independent random-access point for the *sub-bitstream*

Note 1 to entry: A a sub-bitstream IRAP access unit is a sub-bitstream IRAP composition unit.

3.15
view
2D rectangular arrays of *view samples* (3.18) consisting of *attribute* frames and corresponding *geometry* frame representing the projection of a *volumetric frame* onto a surface using *view parameters* (3.16)

3.16
view parameters
parameters of the projection used to generate a *view* (3.15) from a *volumetric frame*, including intrinsic and extrinsic parameters

3.17
view parameters list
listing of one or more *view parameters* (3.16)

3.18
view sample
position on the rectangular frame associated with a *view* (3.15)

3.19
viewing orientation
unit quaternion representing the orientation of a user who is consuming the visual content

3.20
viewing position
triple of x, y, z characterizing the position in the Cartesian coordinates of a user who is consuming the visual content

3.21
viewing space
domain constraints for an intended *viewport* (3.22) rendering

Note 1 to entry: The domain is defined in the 3D global space and related to the *viewing orientation* (3.19); it defines a scale between 0 and 1 for every point in space for a given direction of the *viewport* (3.22), to be used by the application.

3.22
viewport
view (3.15) suitable for display and viewing

4 Abbreviated terms

For the purposes of this document, the abbreviated terms given in ISO/IEC 23090-5 and the following apply.

CSG	constructive solid geometry
ERP	equirectangular projection
HMD	head-mounted display
MIV	MPEG immersive video
OMAF	Omnidirectional media format

5 Conventions

The specifications in ISO/IEC 23090-5:—, Clause 5 and its subclauses apply with the following addition to subclause 5.8:

Cos(x) the trigonometric cosine function operating on an argument x in units of radians

Dot(x, y) dot product function, known also as scalar product function, operating on two vectors x and y

Norm(x) = Sqrt(Norm2(x, x))

Norm2(x) = Abs(Dot(x, x))

Sin(x) the trigonometric sine function operating on an argument x in units of radians

π the ratio of a circle's circumference to its diameter

6 Overall V3C characteristics, decoding operations, and post-decoding processes

The specifications in ISO/IEC 23090-5:—, Clause 6 apply.

7 Bitstream format, partitioning, and scanning processes

7.1 General

The specifications in ISO/IEC 23090-5:—, subclause 7.1 apply.

7.2 V3C bitstream formats

The specifications in ISO/IEC 23090-5:—, subclause 7.2 apply.

7.3 NAL bitstream formats

The specifications in ISO/IEC 23090-5:—, subclause 7.3 apply.

7.4 Partitioning of atlas frames into tiles

The specifications in ISO/IEC 23090-5:—, subclause 7.4 apply.

7.5 Tile partition scanning processes

The specifications in ISO/IEC 23090-5:—, subclause 7.5 apply.

7.6 Mapping of views to V3C components

This subclause describes the concept of views and its mapping to patches in V3C components.

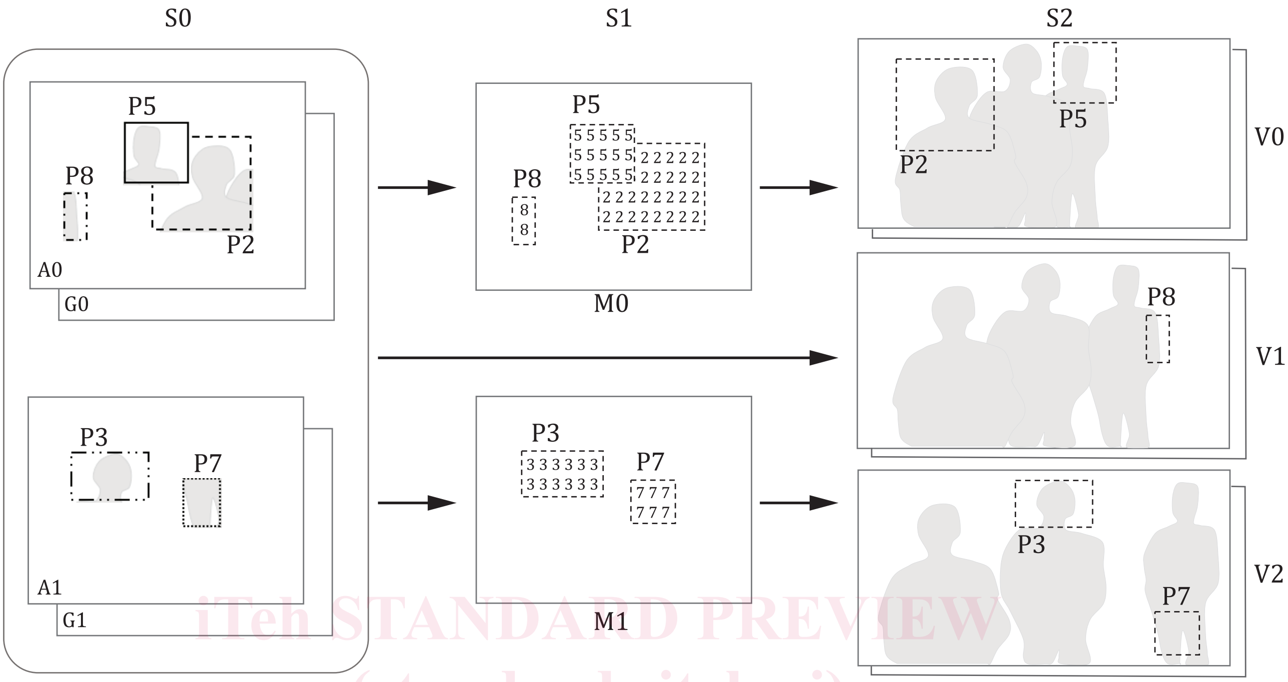
A view represents a field of view of a volumetric frame for a particular view position and orientation. Each view, at a given time instance, may be represented by one 2D frame providing geometry information plus one 2D frame per attribute, providing attribute information, and occupancy information that may either be embedded within geometry or represented explicitly as a 2D frame. Each view may use the equirectangular, perspective, or orthographic projection format. The atlas components of a view use the same projection format.

The volumetric frame and all views each have an associated reference frame. Cartesian coordinates of 3D points can therefore be expressed according to the reference frame of the scene, as represented by the volumetric frame, or according to the reference frame of any view. The camera extrinsic parameters (position and orientation) of the views specify the relations between their reference frames, enabling switching of the 3D coordinate system to represent a 3D point from one reference frame attached to a given view to another reference frame attached to another view.

A coded atlas contains information describing the patches within the atlas. For each patch, a view ID is signalled which identifies which view the patch originated from.

A patch represents a rectangular region of a view, with corresponding regions in all present atlas components: attribute(s), geometry, and occupancy. The size (width and height) of each patch in an atlas is signalled. In this version of the document, the size of a patch is always the same as the corresponding rectangular region in the view texture attribute component, but scaling may optionally be applied to the geometry component or the occupancy component.

Figure 1 shows an illustrative example, in which two atlases contain five patches, which are mapped to three views, with a texture attribute component and a geometry component.



Key

A0-A1 decoded attribute frames for atlas 0 and 1

G0-G1 decoded geometry frames for atlas 0 and 1

M0-M1 maps for atlas 0 and 1

P0-P8 patches

S0 stage 0 where attribute and geometry frames are decoded for each atlas

S1 stage 1 where block to patch mapping is performed

S2 stage 2 where patches are mapped to views

V0-V2 reconstructed views

Figure 1 — Example mapping of 5 patches in 2 atlases to 3 views

7.7 Sources and outputs

The volumetric video source that is represented by the bitstream is a sequence of volumetric frames. Each volumetric frame is represented by one or more view frames, each of which may be represented by a geometry picture, an attribute picture for each attribute, and occupancy information, which may be conveyed in the geometry picture or represented separately.

The outputs of the decoding process are described in subclause 9.1.

The outputs of the non-normative rendering process of Annex H are a sequence of one or more views. The number of views and the associated view parameters may be selected by the application. For example, a single view may be output corresponding to a viewport suitable for display, or a set of views may be output which correspond to the source view parameters.

8 Syntax and semantics

8.1 Method of specifying syntax in tabular form

The specifications in ISO/IEC 23090-5:—, subclause 8.1 apply.

8.2 Specification of syntax functions and descriptors

The specifications in ISO/IEC 23090-5:—, subclause 8.2 apply.

8.3 Syntax in tabular form

8.3.1 General syntax

The specifications in ISO/IEC 23090-5:—, subclause 8.3 apply with the following addition.

An overview of the V3C bitstream structure with MIV extensions is represented in Figure 2.

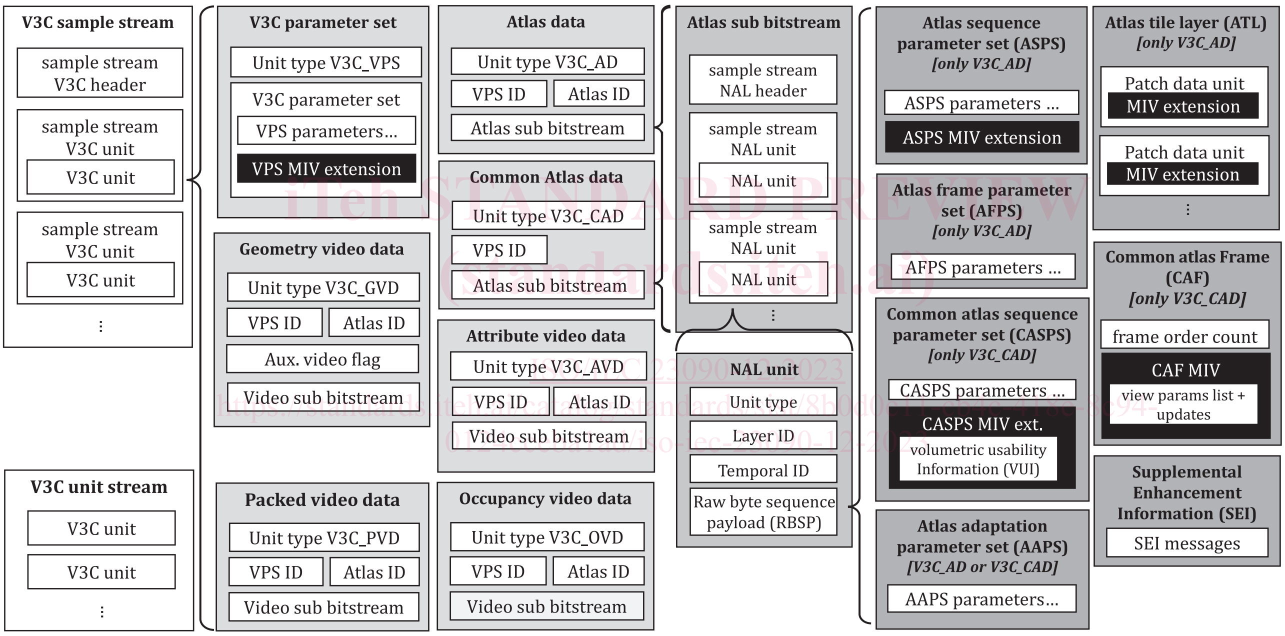


Figure 2 — Overview of V3C bitstream with MIV extensions

8.3.2 V3C unit syntax

The specifications in ISO/IEC 23090-5:—, subclause 8.3.2 apply.

8.3.3 Byte alignment syntax

The specifications in ISO/IEC 23090-5:—, subclause 8.3.3 apply.

8.3.4 V3C parameter set syntax

The specifications in ISO/IEC 23090-5:—, subclause 8.3.4 apply.

8.3.5 NAL unit syntax

The specifications in ISO/IEC 23090-5:—, subclause 8.3.5 apply.

8.3.6 Raw byte sequence payloads, trailing bits, and byte alignment syntax

The specifications in ISO/IEC 23090-5:—, subclause 8.3.6 apply.

8.3.7 Atlas tile data unit syntax

The specifications in ISO/IEC 23090-5:—, subclause 8.3.7 apply.

8.3.8 Supplemental enhancement information message syntax

The specifications in ISO/IEC 23090-5:—, subclause 8.3.8 apply.

8.3.9 V3C MIV extension syntax in tabular form

8.3.9.1 V3C parameter set MIV extension syntax

vps_miv_extension( ) {	**Descriptor**
vme_geometry_scale_enabled_flag	u(1)
vme_embedded_occupancy_enabled_flag	u(1)
if(!vme_embedded_occupancy_enabled_flag)	
vme_occupancy_scale_enabled_flag	u(1)
group_mapping( )	
}	

8.3.9.2 Group mapping syntax

group_mapping( ) {	**Descriptor**
gm_group_count	u(4)
if(gm_group_count > 0)	
for(k = 0; k <= vps_atlas_count_minus1; k++)	
j = vps_atlas_id[k]	
gm_group_id[j]	u(v)
}	

8.3.9.3 Atlas sequence parameter set MIV extension syntax

asps_miv_extension( ) {	**Descriptor**
asme_ancillary_atlas_flag	u(1)
asme_embedded_occupancy_enabled_flag	u(1)
if(asme_embedded_occupancy_enabled_flag)	
asme_depth_occ_threshold_flag	u(1)
asme_geometry_scale_enabled_flag	u(1)
if(asme_geometry_scale_enabled_flag) {	
asme_geometry_scale_factor_x_minus1	ue(v)
asme_geometry_scale_factor_y_minus1	ue(v)
}	
if(!asme_embedded_occupancy_enabled_flag)	
asme_occupancy_scale_enabled_flag	u(1)
if(!asme_embedded_occupancy_enabled_flag && asme_occupancy_scale_enabled_flag) {	

asme_occupancy_scale_factor_x_minus1	ue(v)
asme_occupancy_scale_factor_y_minus1	ue(v)
}	
asme_patch_constant_depth_flag	u(1)
asme_patch_attribute_offset_enabled_flag	u(1)
if(asme_patch_attribute_offset_enabled_flag)	
asme_patch_attribute_offset_bit_depth_minus1	ue(v)
asme_max_entity_id	ue(v)
asme_inpaint_enabled_flag	u(1)
}	

8.3.9.4 Atlas frame parameter set MIV extension syntax

afps_miv_extension( ) {	**Descriptor**
if(!afps_lod_mode_enabled_flag) {	
afme_inpaint_lod_enabled_flag	u(1)
if(afme_inpaint_lod_enabled_flag) {	
afme_inpaint_lod_scale_x_minus1	ue(v)
afme_inpaint_lod_scale_y_idc	ue(v)
}	
}	
}	

8.3.9.5 Common atlas sequence parameter set MIV extension syntax

casps_miv_extension( ) {	**Descriptor**
casme_depth_low_quality_flag	u(1)
casme_depth_quantization_params_present_flag	u(1)
casme_vui_params_present_flag	u(1)
if(casme_vui_params_present_flag)	
vui_parameters( )	
}	

8.3.9.6 Common atlas frame

8.3.9.6.1 MIV extension syntax

caf_miv_extension( ) {	**Descriptor**
if(nal_unit_type == NAL_CAF_IDR) {	
miv_view_params_list( )	
} else {	
came_update_extrinsics_flag	u(1)
came_update_intrinsics_flag	u(1)
if(casme_depth_quantization_params_present_flag)	
came_update_depth_quantization_flag	u(1)
if(came_update_extrinsics_flag)	

miv_view_params_update_extrinsics( )	
if(came_update_intrinsics_flag)	
miv_view_params_update_intrinsics( )	
if(came_update_depth_quantization_flag)	
miv_view_params_update_depth_quantization( )	
}	
}	

8.3.9.6.2 MIV view parameters list syntax

miv_view_params_list( ) {	**Descriptor**
mvp_num_views_minus1	u(16)
mvp_explicit_view_id_flag	u(1)
if(mvp_explicit_view_id_flag)	
for(v = 0; v <= mvp_num_views_minus1; v++)	
mvp_view_id[v]	u(16)
for(v = 0; v <= mvp_num_views_minus1; v++) {	
camera_extrinsics(v)	
mvp_inpaint_flag[v]	u(1)
}	
mvp_intrinsic_params_equal_flag	u(1)
for(v = 0; v <= mvp_intrinsic_params_equal_flag ? 0 : mvp_num_views_minus1; v++)	
camera_intrinsics(v)	
if(casme_depth_quantization_params_present_flag) {	
mvp_depth_quantization_params_equal_flag	u(1)
for(v = 0; v <= mvp_depth_quantization_equal_flag ? 0 : mvp_num_views_minus1; v++)	
depth_quantization(v)	
}	
mvp_pruning_graph_params_present_flag	u(1)
if (mvp_pruning_graph_params_present_flag)	
for(v = 0; v <= mvp_num_views_minus1; v++)	
pruning_parents(v)	
}	

8.3.9.6.3 MIV view parameters update extrinsics syntax

miv_view_params_update_extrinsics( ) {	**Descriptor**
mvpue_num_view_updates_minus1	u(16)
for(i = 0; i <= mvpue_num_view_updates_minus1; i++) {	
mvpue_view_idx[i]	u(16)
camera_extrinsics(mvpue_view_idx[i])	
}	
}	