INTERNATIONAL STANDARD

ISO/IEC 14496-15

Fifth edition
2019-09

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

Technologies de l'information — Codage des objets audiovisuels — Partie 15: Transport de vidéo structurée en unités NAL sur la couche réseau au format ISO de base pour les fichiers médias

Reference number
ISO/IEC 14496-15:2019(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
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).

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 http://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.

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 fifth edition cancels and replaces the fourth edition (ISO/IEC 14496-15:2017), which has been technically revised. It also incorporates Amendments ISO/IEC 14496-15:2017/Amd.1:2018 and ISO/IEC 14496-15:2017/Amd.2:2019.

The main changes compared to the previous edition are as follows:

— additional content incorporated as subclauses 4.11, 4.12, 9.6.4, D.4.3, D.4.4 and D.4.5 and Annex F;

— corrections in Tables 2, 3 and 6 and subclause A.1;

— deletion of subclause 5.4.10;

— minor editorial changes to align the document with the drafting rules in ISO/IEC Directives Part 2.

A list of all parts in the ISO/IEC 14496 series can be found on the ISO website.

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.

Introduction

This document defines a storage format based on, and compatible with, the ISO base media file format (ISO/IEC 14496-12), which is used by the MP4 file format (ISO/IEC 14496-14) and the motion JPEG 2000 file format (ISO/IEC 15444-3) among others. This document enables video streams formatted as network adaptation layer units (NAL units) to

a) be used in conjunction with other media streams, such as audio,

b) be used in an MPEG-4 systems environment, if desired,

c) be formatted for delivery by a streaming server, using hint tracks, and

d) inherit all the use cases and features of the ISO base media file format on which MP4 and MJ2 are based.

This document can be used as a standalone specification; it specifies how NAL unit structured video content is stored in an ISO base media file format compliant format. However, it is normally used in the context of a specification, such as the MP4 file format, derived from the ISO base media file format, that permits the use of NAL unit structured video such as AVC (ISO/IEC 14496-10) video and high efficiency video coding (HEVC, ISO/IEC 23008-2) video.

The ISO base media file format is becoming increasingly common as a general-purpose media container format for the exchange of digital media, and its use in this context should accelerate both adoption and interoperability.

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 patents.

ISO and IEC take no position concerning the evidence, validity and scope of these patent rights. The holders of these patent rights have assured ISO and IEC that they are willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statements of the holders of these patent rights are registered with ISO and IEC. Informstion may be obtained from:

Fraunhofer-Gesellschaft
Hansastr. 27c
80686 München
Germany

Qualcomm Incorporated
5775 Morehouse Drive,
San Diego, CA 92121
USA

Nokia Corporation
PO Box 86
FIN-24101 Salo
Finland

Huawei Technologies Co., Ltd.
Bantian Longgang District,
Shenzhen 518129
China

TDVision Systems, Inc.
8001 Irvine Center Drive, Suite 400
Irvine, CA 92618
USA

Telefonaktiebolaget LM Ericsson (publ)
Ericsson AB,
SE-164 80 Stockholm
Sweden

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

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

1 Scope

This document specifies the storage format for streams of video that is structured as NAL units, such as AVC (ISO/IEC 14496-10) and HEVC (ISO/IEC 23008-2) video streams.

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 14496-10:2014, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ISO/IEC 14496-12:2015, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 23008-2:2017, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

3 Terms, definitions and abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-10, ISO/IEC 23008-2 and the following apply.

3.1.1
3D-AVC NAL unit
3D-AVC VCL NAL unit
NAL unit with type 21 with avc_3d_extension_flag equal to 1

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex J.

3.1.2
aggregator
in-stream structure (3.1.11) using a NAL unit header for grouping of NAL units belonging to the same sample

3.1.3
AVC base layer
maximum subset of a bitstream that is AVC compatible

Note 1 to entry: This can also be expressed as a bitstream not using any of the functionality of ISO/IEC 14496-10:2014 Annex G, Annex H, Annex I, or Annex J.

Note 2 to entry: The AVC base layer is represented by AVC VCL NAL units and associated non-VCL NAL units.

Note 3 to entry: The AVC base layer itself can be a temporal scalable bitstream.

3.1.4
AVC NAL unit
AVC VCL NAL unit (3.1.5) and its associated non-VCL NAL units in a bitstream

3.1.5
AVC VCL NAL unit
NAL unit with type 1 to 5 (inclusive) Note 1 to entry: As specified in ISO/IEC 14496-10.

3.1.6
complete subset
minimal set of tracks that contain all the information in the original bitstream

3.1.7
cropped frame dimensions
width and height of the decoded frame after applying the output cropping parameters specified by the active SPS

3.1.8
extraction path
set of operations on the original bitstream, each yielding a subset bitstream, ordered such that the complete bitstream is first in the set, and the base layer is last, and all the bitstreams are in decreasing complexity (along one of the scalability axes, such as resolution), and where every bitstream is a valid operating point

Note 1 to entry: An extraction path can be represented by the values of priority_id in the NAL unit headers. Alternatively, an extraction path can be represented by the run of tiers or by a set of hierarchically dependent tracks.

3.1.9
extractor
in-stream structure (3.1.11) using a NAL unit header for extraction of data from other tracks

Note 1 to entry: Extractors contain instructions on how to extract data from other tracks. Logically an Extractor can be seen as a pointer to data. While reading a track containing Extractors, the Extractor is replaced by the data it is pointing to.

3.1.10
implicit reconstruction
reconstruction of a stream of access units from two or more tracks not using extractors

3.1.11
in-stream structure
structure residing within sample data

3.1.12
layer set
set of layers represented within a bitstream created from another bitstream by operation of the sub-bitstream extraction process

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.13
MVC NAL unit
MVC VCL NAL unit (3.1.14) and its associated non-VCL NAL units in an MVC streamNote 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex H.

3.1.14
MVC VCL NAL unit
NAL unit with type 20, and NAL units with type 14 when the immediately following NAL units are AVC VCL NAL units

Note 1 to entry: As specified in ISO/IEC 14496-10.

Note 2 to entry: MVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.15
MVC+D depth NAL unit
MVC+D depth VCL NAL unit
NAL unit with type 21 containing a coded slice extension for a depth view component Note 1 to entry: As specified in ISO/IEC 14496-10:2014 Annex I.

3.1.16
MVD NAL unit
MVD VCL NAL unit
NAL unit with type 21, containing a coded slice extension for a depth view component coded with MVC+D or 3D-AVC, or a 3D-AVC texture view component

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex I or J.

3.1.17
NAL-unit-like structure
data structure that is similar to NAL units in the sense that it also has a NAL unit header and a payload, with a difference that the payload might not follow the start code emulation prevention mechanism required for the NAL unit syntax

Note 1 to entry: As specified in ISO/IEC 14496-10 or ISO/IEC 23008-2.

3.1.18
natively present
not included in an *aggregator* (3.1.2) or an *extractor* (3.1.9)

Note 1 to entry: Data referred to by (hence not included in) an aggregator is considered as natively present. Data included in an aggregator is not considered as natively present.

3.1.19
operating point
<AVC and SVC elementary stream> independently decodable subset of a layered bitstream

Note 1 to entry: Each operating point consists of all the data needed to decode this particular bitstream subset.

Note 2 to entry: In an SVC stream an operating point represents a particular spatial resolution, temporal resolution, and quality, and can be represented either by (i) specific values of DTQ (dependency_id, temporal_id and quality_id) or (ii) specific values of P (priority_id) or (iii) combinations of them (e.g. PDTQ). Note that the usage of priority_id is defined by the application. In an SVC file a track represents one or more operating points. Within a track tiers can be used to define multiple operating points.

Note 3 to entry: The bitstream subset of an MVC or MVD operating point represents a particular set of target output views at a particular temporal resolution and consists of all the data needed to decode this particular bitstream subset. In MVD each target output view in the bitstream subset of an MVD operating point can contain a texture view, a depth view or both.

Note 4 to entry: An operating point is referred to as an operation point in ISO/IEC 14496-10:2014, Annex H or an output operation point in ISO/IEC 23008-2.

3.1.20
operating point
<HEVC, layered HEVC and tiled HEVC elementary stream> independently decodable subset of a layered bitstream, where one or more layers in the set of layers are indicated to be output layers

Note 1 to entry: Each operating point consists of all the data needed to decode this particular bitstream subset.

Note 2 to entry: In an SVC stream an operating point represents a particular spatial resolution, temporal resolution, and quality, and can be represented either by (i) specific values of DTQ (dependency_id, temporal_id and quality_id) or (ii) specific values of P (priority_id) or (iii) combinations of them (e.g. PDTQ). Note that the usage of priority_id is defined by the application. In an SVC file a track represents one or more operating points. Within a track tiers can be used to define multiple operating points.

Note 3 to entry: The bitstream subset of an MVC or MVD operating point represents a particular set of target output views at a particular temporal resolution and consists of all the data needed to decode this particular bitstream subset. In MVD each target output view in the bitstream subset of an MVD operating point can contain a texture view, a depth view or both.

Note 4 to entry: An operating point is referred to as an operation point in ISO/IEC 14496-10:2014, Annex H or an output operation point in ISO/IEC 23008-2.

3.1.21
output layer set
set of layers consisting of the layers of one of the specified *layer sets* (3.1.12), where one or more layers in the set of layers are indicated to be output layers

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.22
parameter set
video parameter set, sequence parameter set or picture parameter set

Note 1 to entry: As defined in the applicable video standard (e.g. ISO/IEC 14496-10 or ISO/IEC 23008-2).

Note 2 to entry: This term is used to refer to all types of parameter sets.

3.1.23
parameter set elementary stream
elementary stream containing samples made up of only sequence and picture parameter set NAL units synchronized with the *video elementary stream* (3.1.40)

3.1.24
picture unit
set of VCL NAL units and their associated non-VCL NAL units

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.25
prefix NAL unit
NAL units with type 14

Note 1 to entry: As specified in ISO/IEC 14496-10.

Note 2 to entry: Prefix NAL units provide scalability information about AVC VCL NAL units and filler data NAL units. Prefix NAL units do not affect the decoding process of a legacy AVC decoder. The behaviour of a legacy AVC file reader as a response to prefix NAL units is undefined.

3.1.26
reference layer
layer that is indicated as possibly needed for decoding of another layer

Note 1 to entry: As specified in ISO/IEC 23008-2 and as specified by the `'oinf'` sample group defined in subclause 9.6.2.

3.1.27
scalable layer
layer
<AVC and SVC elementary stream> set of VCL NAL units with the same values of dependency_id, quality_id, and temporal_id, and the associated non-VCL NAL units

Note 1 to entry: As specified in ISO/IEC 14496-10.

Note 2 to entry: A scalable layer with any of dependency_id, quality_id, and temporal_id not equal to 0 enhances the video by one or more scalability levels in at least one direction (temporal, quality or spatial resolution).

Note 3 to entry: SVC uses a "layered" encoder design that results in a bitstream representing "coding layers". In some publications the 'base layer' is the first quality layer of a specific coding layer. In some publications the base layer is the scalable layer with the lowest priority. The SVC file format uses "scalable layer" or "layer" in a general way for describing nested bitstreams (using terms like AVC base layer or SVC enhancement layer).

3.1.28
scalable layer
layer
<HEVC, layered HEVC and tiled HEVC elementary stream> set of VCL NAL units with the same value of nuh_layer_id and the associated non-VCL NAL units

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.29
scalable layer representation
bitstream subset that is required for decoding the *scalable layer* (3.1.27/3.1.28), consisting of the scalable layer itself and all the scalable layers on which the scalable layer depends

Note 1 to entry: A scalable layer representation is also referred to as the representation of the scalable layer.

3.1.30
sub-picture
proper subset of coded slices of a layer representation

3.1.31
sub-picture tier
tier that consists of *sub-pictures* (3.1.30)

Note 1 to entry: Any coded slice that is not included in the tier representation of a sub-picture tier is not to be referred to in inter prediction or inter-layer prediction for decoding of the sub-picture tier.

3.1.32
sub-layer
set of VCL NAL units with a particular value of TemporalId and the associated non-VCL NAL units

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.33
SVC enhancement layer
layer that specifies a part of a scalable bitstream that enhances the video

Note 1 to entry: An SVC enhancement layer is represented by SVC VCL NAL units and the associated non-VCL NAL units and SEI messages.

Note 2 to entry: Usually an SVC enhancement layer represents a spatial or coarse-grain scalability (CGS) coding layer (identified by a specific value of dependency_id).

3.1.34
SVC NAL unit
SVC VCL NAL unit (3.1.36) and its associated non-VCL NAL units in an *SVC stream* (3.1.35)

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex G.

3.1.35
SVC stream
bitstream represented by the operating point for which dependency_id is equal to mDid, temporal_id is the greatest temporal_id value among mOpSet, and quality_id is the greatest quality_id value among mOpSet, where the greatest value of dependency_id of all the operating points represented by DTQ (dependency_id, temporal_id and quality_id) combinations is equal to mDid, and the set of all the operating points with dependency_id equal to mDid is mOpSet

Note 1 to entry: The term "SVC stream" is referenced by 'decoding/accessing the entire stream' in this document. There can be NAL units that are not required for decoding this operating point.

3.1.36
SVC VCL NAL unit
NAL unit with type 20, and NAL units with type 14 when the immediately following NAL units are AVC VCL NAL units

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex G.

Note 2 to entry: SVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.37
temporal layer representation
representation of a temporal layer
temporal layer and all lower temporal layers

3.1.38
tier
set of operating points within a track, providing information about the operating points and instructions on how to access the corresponding bitstream portions (using maps and groups)

Note 1 to entry: In SVC file format a tier represents one or more scalable layers of an SVC bitstream. In the context of ISO/IEC 23008-2 video, the term tier is used to represent a part of the interoperability point representation consisting of profile, tier, and level. Readers should not be confused about these two different meanings of the word "tier".

Note 2 to entry: The term "tier" is used in SVC file format to avoid confusion with the frequently used term layer. A tier represents a subset of a track and represents an operating point of an SVC bitstream. Tiers in a track subset the entire track, no matter whether the track references another track by extractors.

Note 3 to entry: An MVC or MVD tier represents a particular set of temporal subsets of a particular set of views.

3.1.39
tier representation
representation of the tier
bitstream subset that is required for decoding the tier, consisting of the tier itself and all the tiers on which the tier depends

3.1.40
video elementary stream
elementary stream containing access units made up of NAL units for coded picture data

3.1.41
video stream
self-contained independently decodable video bitstream

3.1.42
virtual base view
AVC compatible representation of an independently coded non-base view

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex H.

Note 2 to entry: The virtual base view of an independently coded non-base view is created according to the process specified in ISO/IEC 14496-10:2014, H.8.5.5. Samples containing data units of an independently coded non-base view and samples of the virtual base view are aligned by decoding times.

3.2 Abbreviated terms

3D-AVC	three-dimensional advanced video coding (refers to ISO/IEC 14496-10:2014, Annex J)
3D-HEVC	three-dimensional high efficiency video coding (refers to ISO/IEC 23008-2:2017, Annex I)
A3D	three-dimensional advanced video coding (refers to ISO/IEC 14496-10:2014, Annex G) NOTE 1 The abbreviation A3D is used in terminology related to syntax elements and structures, whereas the abbreviation 3D-AVC is used otherwise.
AVC	advanced video coding [where contrasted with SVC, MVC, or MVD in this document, this term refers to ISO/IEC 14496-10, excluding its Annex G (Scalable Video Coding), Annex H (Multiview Video Coding), Annex I (Multiview and Depth Video Coding), and Annex J (Multiview and Depth Video with Enhanced Non-Base View Coding)]
BLA	broken link access
CRA	clean random access
CTU	coding tree unit
HEVC	high efficiency video coding
FF	file format
HRD	hypothetical reference decoder
IDR	instantaneous decoding refresh
IRAP	intra random access point
L-HEVC	layered high efficiency video coding
MVC	multiview video coding (refers to ISO/IEC 14496-10:2014, Annex H)
MVCD	multiview video coding plus depth (refers to ISO/IEC 14496-10:2014, Annex I)
MVC+D	multiview video coding plus depth (refers to ISO/IEC 14496-10:2014, Annex I) NOTE 2 The abbreviation MVCD is used in terminology related to syntax elements and structures, whereas the abbreviation MVC+D is used otherwise.
MV-HEVC	multiview high efficiency video coding (refers to ISO/IEC 23008-2:2017, Annex G)
MVD	multiview video coding plus depth (refers to ISO/IEC 14496-10:2014, Annexes I and J)