INTERNATIONAL STANDARD

ISO/IEC 23008-2

Fourth edition
2020-08

Information technology — High efficiency coding and media delivery in heterogeneous environments —

Part 2: High efficiency video coding

Technologies de l'information — Codage à haute efficacité et livraison des medias dans des environnements hétérogènes —
Partie 2: Codage vidéo à haute efficacité

Reference number
ISO/IEC 23008-2:2020(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).

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).

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 the following URL: www.iso.org/iso/foreword.html.

This document was prepared by ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*, in collaboration with ITU-T. The technically identical text is published as Rec. ITU-T H.265.

This fourth edition cancels and replaces the third edition (ISO/IEC 23008-2:2017), which has been technically revised. It also incorporates the amendments ISO/IEC 23008-2:2017/Amd 1:2018, ISO/IEC 23008-2:2017/Amd 2:2018, and ISO/IEC 23008-2:2017/Amd 3:2018).

The main changes compared to the previous edition are:

— the specification of two additional profiles (the Monochrome 10 and Main 10 Still Picture profiles);

— the specification of additional supplemental enhancement information (SEI) messages and additional colour-related video usability information codepoint identifiers;

— corrections to various minor defects.

A list of all parts in the ISO/IEC 23008 series can be found on the ISO website.

Introduction

As the costs for both processing power and memory have reduced, network support for coded video data has diversified, and advances in video coding technology have progressed. The need has arisen for an industry standard for compressed video representation with substantially increased coding efficiency and enhanced robustness to network environments. Toward these ends, the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) formed a Joint Collaborative Team on Video Coding (JCT-VC) in 2010 and a Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) in 2012 for development of a new Recommendation | International Standard. This document was developed in the JCT-VC and the JCT-3V.

Purpose

This document was developed in response to the growing need for higher compression of moving pictures for various applications such as videoconferencing, digital storage media, television broadcasting, internet streaming, and communications. It is also designed to enable the use of the coded video representation in a flexible manner for a wide variety of network environments as well as to enable the use of multi-core parallel encoding and decoding devices. The use of this document allows motion video to be manipulated as a form of computer data and to be stored on various storage media, transmitted and received over existing and future networks and distributed on existing and future broadcasting channels. Supports for higher bit depths and enhanced chroma formats, including the use of full-resolution chroma are provided. Support for scalability enables video transmission on networks with varying transmission conditions and other scenarios involving multiple bit rate services. Support for multiview enables representation of video content with multiple camera views and optional auxiliary information. Support for 3D enables joint representation of video content and depth information with multiple camera views.

Applications

This document is designed to cover a broad range of applications for video content including but not limited to the following:

— broadcast (cable TV on optical networks / copper, satellite, terrestrial, etc.);

— camcorders;

— content production and distribution;

— digital cinema;

— home cinema;

— internet streaming, download and play;

— medical imaging;

— mobile streaming, broadcast and communications;

— real-time conversational services (videoconferencing, videophone, telepresence, etc.);

— remote video surveillance;

— storage media (optical disks, digital video tape recorder, etc.);

— wireless display.

Publication and versions of this document

This document has been jointly developed by ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). It is published as technically-aligned twin text in both ITU-T and ISO/IEC. As the basis text has been drafted to become both an ITU-T Recommendation and an ISO/IEC International Standard, the term "Specification" (with capitalization to indicate that it refers to the whole of the text) is used herein when the text refers to itself.

This is the fifth version of this document and the fourth edition published by ISO/IEC.

Rec. ITU-T H.265 | ISO/IEC 23008-2 version 1 refers to the first approved version of this document. The first edition published by ISO/IEC as ISO/IEC 23008-2:2013 corresponded to the first version.

Rec. ITU-T H.265 | ISO/IEC 23008-2 version 2 refers to the integrated text additionally containing format range extensions, scalability extensions, multiview extensions, additional supplement enhancement information, and corrections to various minor defects in the prior content of the specification. The second edition published by ISO/IEC as ISO/IEC 23008-2:2015 corresponded to the second version.

Rec. ITU-T H.265 | ISO/IEC 23008-2 version 3 refers to the integrated text additionally containing 3D extensions, additional supplement enhancement information, and corrections to various minor defects in the prior content of the specification.

Rec. ITU-T H.265 | ISO/IEC 23008-2 version 4 refers to the integrated text additionally containing screen content coding extensions profiles, scalable range extensions profiles, additional high throughput profiles, additional supplement enhancement information, additional colour representation identifiers, and corrections to various minor defects in the prior content of the specification. The third edition published by ISO/IEC as ISO/IEC 23008-2:2017 corresponded to the fourth version.

Rec. ITU-T H.265 | ISO/IEC 23008-2 version 5 refers to the integrated text additionally containing additional SEI messages that include omnidirectional video specific SEI messages, a Monochrome 10 profile, a Main 10 Still Picture profile, and corrections to various minor defects in the prior content of the specification.

Rec. ITU-T H.265 | ISO/IEC 23008-2 version 6 refers to the integrated text additionally containing additional SEI messages for SEI manifest and SEI prefix, and corrections to various minor defects in the prior content of the specification.

Rec. ITU-T H.265 | ISO/IEC 23008-2 version 7 (the current version) refers to the integrated text additionally containing the fisheye video information SEI message and the annotated regions SEI message, and corrections to various minor defects in the prior content of the specification. The fourth edition published by ISO/IEC as ISO/IEC 23008-2:2020 corresponds to the fifth version.

Profiles, tiers, and levels

This document is designed to be generic in the sense that it serves a wide range of applications, bit rates, resolutions, qualities, and services. Applications should cover, among other things, digital storage media, television broadcasting, and real-time communications. In the course of creating this document, various requirements from typical applications have been considered, necessary algorithmic elements have been developed, and these have been integrated into a single syntax. Hence, this document will facilitate video data interchange among different applications.

Considering the practicality of implementing the full syntax of this document, however, a limited number of subsets of the syntax are also stipulated by means of "profiles", "tiers", and "levels". These and other related terms are formally defined in Clause 3.

A “profile” is a subset of the entire bitstream syntax that is specified in this document. Within the bounds imposed by the syntax of a given profile it is still possible to require a very large variation in the performance of encoders and decoders depending upon the values taken by syntax elements in the bitstream such as the specified size of the decoded pictures. In many applications, it is currently neither practical nor economical to implement a decoder capable of dealing with all hypothetical uses of the syntax within a particular profile.

In order to deal with this problem, "tiers" and "levels" are specified within each profile. A level of a tier is a specified set of constraints imposed on values of the syntax elements in the bitstream. These constraints may be simple limits on values. Alternatively they may take the form of constraints on arithmetic combinations of values (e.g. picture width multiplied by picture height multiplied by number of pictures decoded per second). A level specified for a lower tier is more constrained than a level specified for a higher tier.

Coded video content conforming to this document uses a common syntax. In order to achieve a subset of the complete syntax, flags, parameters, and other syntax elements are included in the bitstream that signal the presence or absence of syntactic elements that occur later in the bitstream.

Overview of the design characteristics

The coded representation specified in the syntax is designed to enable a high compression capability for a desired image or video quality. The algorithm is typically not lossless, as the exact source sample values are typically not preserved through the encoding and decoding processes. A number of techniques may be used to achieve highly efficient compression. Encoding algorithms (not specified in this Recommendation | International Standard) may select between inter and intra coding for block-shaped regions of each picture. Inter coding uses motion vectors for block-based inter prediction to exploit temporal statistical dependencies between different pictures. Intra coding uses various spatial prediction modes to exploit spatial statistical dependencies in the source signal for a single picture. Motion vectors and intra prediction modes may be specified for a variety of block sizes in the picture. The prediction residual may then be further compressed using a transform to remove spatial correlation inside the transform block before it is quantized, producing a possibly irreversible process that typically discards less important visual information while forming a close approximation to the source samples. Finally, the motion vectors or intra prediction modes may also be further compressed using a variety of prediction mechanisms, and, after prediction, are combined with the quantized transform coefficient information and encoded using arithmetic coding.

How to read this document

It is suggested that the reader starts with Clause 1 (Scope) and moves on to Clause 3 (Terms and definitions). Clause 6 should be read for the geometrical relationship of the source, input, and output of the decoder. Clause 7 (Syntax and semantics) specifies the order to parse syntax elements from the bitstream. See 7.1 to 7.3 for syntactical order and see 7.4 for semantics; e.g. the scope, restrictions, and conditions that are imposed on the syntax elements. The actual parsing for most syntax elements is specified in Clause 9 (Parsing process). Clause 10 (Sub-bitstream extraction process) specifies the sub-bitstream extraction process. Finally, Clause 8 (Decoding process) specifies how the syntax elements are mapped into decoded samples. Throughout reading this document, the reader should refer to Clauses 2 (Normative references), 4 (Abbreviations), and 5 (Conventions) as needed. Annexes A through I also form an integral part of this document.

Annex A specifies profiles each being tailored to certain application domains, and defines the so-called tiers and levels of the profiles. Annex B specifies syntax and semantics of a byte stream format for delivery of coded video as an ordered stream of bytes. Annex C specifies the hypothetical reference decoder, bitstream conformance, decoder conformance, and the use of the hypothetical reference decoder to check bitstream and decoder conformance. Annex D specifies syntax and semantics for supplemental

enhancement information message payloads. Annex E specifies syntax and semantics of the video usability information parameters of the sequence parameter set. Annex F specifies general multi-layer support for bitstreams and decoders. Annex G contains support for multiview coding. Annex H contains support for scalability. Annex I contains support for 3D coding.

Throughout this document, statements appearing with the preamble "NOTE" are informative and are not an integral part of this document.

Patent rights identified

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. 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 — High efficiency coding and media delivery in heterogeneous environments —

Part 2:
High efficiency video coding

1 Scope

This document specifies high efficiency video coding.

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 11664-1, *Colorimetry — Part 1: CIE standard colorimetric observers*

ISO 12232, *Photography — Digital still cameras — Determination of exposure index, ISO speed ratings, standard output sensitivity, and recommended exposure index*

ISO/IEC 10646, *Information technology — Universal Coded Character Set (UCS)*

ISO/IEC 11578, *Information technology — Open Systems Interconnection — Remote Procedure Call (RPC)*

ISO/IEC 23001-11, *Information technology — MPEG systems technologies — Part 11: Energy-efficient media consumption (green metadata)*

IETF RFC 1321, *The MD5 Message-Digest Algorithm*

IETF RFC 5646, *Tags for Identifying Languages*

Recommendation ITU-T T.35, *Procedure for the allocation of ITU-T defined codes for non-standard facilities*

CEN/TR 13233, *Advanced technical ceramics — Notations and symbols*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in CEN/TR 13233 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
access unit
set of *NAL units* (3.87) that are associated with each other according to a specified classification rule, are consecutive in *decoding order* (3.44), and contain exactly one *coded picture* (3.25) with nuh_layer_id equal to 0

Note 1 to entry: In addition to containing the *VCL NAL units* (3.181) of the coded picture with nuh_layer_id equal to 0, an access unit may also contain non-VCL NAL units. The decoding of an access unit with the decoding process specified in Clause 8 always results in a decoded picture with nuh_layer_id equal to 0.

Note 2 to entry: An access unit is defined differently in Annex F and does not need to contain a coded picture with nuh_layer_id equal to 0.

3.2
AC transform coefficient
transform coefficient (3.175) for which the *frequency index* (3.58) in at least one of the two dimensions is non-zero

3.3
associated IRAP picture
previous *IRAP picture* (3.71) in *decoding order* (3.44) (when present)

3.4
associated non-VCL NAL unit
non-VCL NAL unit (3.91) (when present) for a *VCL NAL unit* (3.180) where the VCL NAL unit is the *associated VCL NAL unit* (3.5) of the non-VCL NAL unit

3.5
associated VCL NAL unit
preceding *VCL NAL unit* (3.180) in *decoding order* (3.44) for a *non-VCL NAL unit* (3.91) with nal_unit_type equal to EOS_NUT, EOB_NUT, FD_NUT, or SUFFIX_SEI_NUT, or in the ranges of RSV_NVCL45..RSV_NVCL47 or UNSPEC56..UNSPEC63; or otherwise, the next VCL NAL unit in decoding order

3.6
azimuth circle
circle on a sphere connecting all points with the same azimuth value

Note 1 to entry: An azimuth circle is always a great circle like a longitude line on the earth.

3.7
base layer
layer in which all *NAL units* (3.87) have nuh_layer_id equal to 0

3.8
bin
one bit of a *bin string* (3.11)

3.9
binarization
set of *bin strings* (3.11) for all possible values of a *syntax element* (3.162)

3.10
binarization process
unique mapping process of all possible values of a *syntax element* (3.162) onto a set of *bin strings* (3.11)

3.11
bin string
intermediate binary representation of values of *syntax elements* (3.162) from the *binarization* (3.9) of the syntax element

3.12
bi-predictive slice
B slice
slice (3.143) that is decoded using *intra prediction* (3.69) or using *inter prediction* (3.67) with at most two *motion vectors* (3.86) and *reference indices* (3.127) to predict the sample values of each *block* (3.14)

3.13
bitstream
sequence of bits, in the form of a *NAL unit stream* (3.88) or a *byte stream* (3.21), that forms the representation of *coded pictures* (3.25) and associated data forming one or more *CVSs* (3.30)

3.14
block
MxN (M-column by N-row) array of samples or an MxN array of *transform coefficients* (3.175)

3.15
broken link
location in a *bitstream* (3.13) at which it is indicated that some subsequent *pictures* (3.99) in *decoding order* (3.44) may contain serious visual artefacts due to unspecified operations performed in the generation of the bitstream

3.16
broken link access access unit
BLA access unit
access unit (3.1) in which the *coded picture* (3.25) with nuh_layer_id equal to 0 is a *BLA picture* (3.17)

3.17
broken link access picture
BLA picture
IRAP picture (3.71) for which each *VCL NAL unit* (3.181) has nal_unit_type equal to BLA_W_LP, BLA_W_RADL, or BLA_N_LP

Note 1 to entry: A BLA picture does not refer to any pictures other than itself for *inter prediction* (3.67) in its *decoding process* (3.45), and may be the first picture in the *bitstream* (3.13) in decoding order, or may appear later in the bitstream. Each BLA picture begins a new *CVS* (3.30), and has the same effect on the decoding process as an *IDR picture* (3.65). However, a BLA picture contains *syntax elements* (3.162) that specify a non-empty *RPS* (3.132). When a BLA picture for which each VCL NAL unit has nal_unit_type equal to BLA_W_LP, it may have associated RASL pictures, which are not output by the decoder and may not be decodable, as they may contain references to pictures that are not present in the bitstream. When a BLA picture for which each VCL NAL unit has nal_unit_type equal to BLA_W_LP, it may also have associated RADL pictures, which are specified to be decoded. When a BLA picture for which each VCL NAL unit has nal_unit_type equal to BLA_W_RADL, it does not have associated RASL pictures but may have associated RADL pictures. When a BLA picture for which each VCL NAL unit has nal_unit_type equal to BLA_N_LP, it does not have any associated leading pictures.

3.18
buffering period
set of *access units* (3.1) starting with an access unit that contains a buffering period SEI message and containing all subsequent access units in *decoding order* (3.44) up to but not including the next access unit (when present) that contains a buffering period SEI message

3.19
byte
sequence of 8 bits, within which, when written or read as a sequence of bit values, the left-most and right-most bits represent the most and least significant bits, respectively

3.20
byte-aligned
position in a *bitstream* (3.13) is byte-aligned when the position is an integer multiple of 8 bits from the position of the first bit in the bitstream and a bit or *byte* (3.19) or *syntax element* (3.162) is said to be byte-aligned when the position at which it appears in *a* bitstream is byte-aligned