INTERNATIONAL STANDARD

ISO/IEC 23008-2

Third edition
2017-10

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 23008-2:2017(E)

COPYRIGHT PROTECTED DOCUMENT

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

Foreword xi

Introduction xii

1 Scope 1

2 Normative references 1

3 Terms and definitions 1

4 Abbreviated terms 19

5 Conventions 20

5.1 General 20

5.2 Arithmetic operators 21

5.3 Logical operators 21

5.4 Relational operators 21

5.5 Bit-wise operators 21

5.6 Assignment operators 22

5.7 Range notation 22

5.8 Mathematical functions 22

5.9 Order of operation precedence 23

5.10 Variables, syntax elements, and tables 24

5.11 Text description of logical operations 25

5.12 Processes 26

6 Bitstream and picture formats, partitionings, scanning processes, and neighbouring relationships 26

6.1 Bitstream formats 26

6.2 Source, decoded, and output picture formats 27

6.3 Partitioning of pictures, slices, slice segments, tiles, coding tree units, and coding tree blocks 29

6.3.1 Partitioning of pictures into slices, slice segments, and tiles 29

6.3.2 Block and quadtree structures 31

6.3.3 Spatial or component-wise partitionings 31

6.4 Availability processes 32
6.4.1 Derivation process for z-scan order block availability 32
6.4.2 Derivation process for prediction block availability 33
6.5 Scanning processes 34
6.5.1 Coding tree block raster and tile scanning conversion process 34
6.5.2 Z-scan order array initialization process 36
6.5.3 Up-right diagonal scan order array initialization process 36
6.5.4 Horizontal scan order array initialization process 37
6.5.5 Vertical scan order array initialization process 37
6.5.6 Traverse scan order array initialization process 37
7 Syntax and semantics 38
7.1 Method of specifying syntax in tabular form 38
7.2 Specification of syntax functions and descriptors 39
7.3 Syntax in tabular form 41
7.3.1 NAL unit syntax 41
7.3.2 Raw byte sequence payloads, trailing bits, and byte alignment syntax 42
7.3.3 Profile, tier and level syntax 51
7.3.4 Scaling list data syntax 53
7.3.5 Supplemental enhancement information message syntax 54
7.3.6 Slice segment header syntax 54
7.3.7 Short-term reference picture set syntax 59
7.3.8 Slice segment data syntax 60
7.4 Semantics 75
7.4.1 General 75
7.4.2 NAL unit semantics 75
7.4.3 Raw byte sequence payloads, trailing bits, and byte alignment semantics 85
7.4.4 Profile, tier, and level semantics 106
7.4.5 Scaling list data semantics 109
7.4.6 Supplemental enhancement information message semantics 111

7.4.7 Slice segment header semantics 112
7.4.8 Short-term reference picture set semantics 120
7.4.9 Slice segment data semantics 123
8 Decoding process 138
8.1 General decoding process 138
8.1.1 General 138
8.1.2 CVSG decoding process 139
8.1.3 Decoding process for a coded picture with nuh_layer_id equal to 0 139
8.2 NAL unit decoding process 141
8.3 Slice decoding process 141
8.3.1 Decoding process for picture order count 141
8.3.2 Decoding process for reference picture set 142
8.3.3 Decoding process for generating unavailable reference pictures 148
8.3.4 Decoding process for reference picture lists construction 149
8.3.5 Decoding process for collocated picture and no backward prediction flag 150
8.4 Decoding process for coding units coded in intra prediction mode 150
8.4.1 General decoding process for coding units coded in intra prediction mode 150
8.4.2 Derivation process for luma intra prediction mode 155
8.4.3 Derivation process for chroma intra prediction mode 157
8.4.4 Decoding process for intra blocks 158
8.5 Decoding process for coding units coded in inter prediction mode 170
8.5.1 General decoding process for coding units coded in inter prediction mode 170
8.5.2 Inter prediction process 171
8.5.3 Decoding process for prediction units in inter prediction mode 174
8.5.4 Decoding process for the residual signal of coding units coded in inter prediction mode 206
8.6 Scaling, transformation and array construction process prior to deblocking filter process 210
8.6.1 Derivation process for quantization parameters 210
8.6.2 Scaling and transformation process 212

8.6.3 Scaling process for transform coefficients 214
8.6.4 Transformation process for scaled transform coefficients 215
8.6.5 Residual modification process for blocks using a transform bypass 218
8.6.6 Residual modification process for transform blocks using cross-component prediction 218
8.6.7 Picture construction process prior to in-loop filter process 219
8.6.8 Residual modification process for blocks using adaptive colour transform 219
8.7 In-loop filter process 222
8.7.1 General 222
8.7.2 Deblocking filter process 222
8.7.3 Sample adaptive offset process 239
9 Parsing process 242
9.1 General 242
9.2 Parsing process for 0-th order Exp-Golomb codes 242
9.2.1 General 242
9.2.2 Mapping process for signed Exp-Golomb codes 244
9.3 CABAC parsing process for slice segment data 244
9.3.1 General 244
9.3.2 Initialization process 247
9.3.3 Binarization process 260
9.3.4 Decoding process flow 269
9.3.5 Arithmetic encoding process (informative) 285

10 Sub-bitstream extraction process 291

Annex A (normative) Profiles, tiers and levels 293

A.1 Overview of profiles, tiers and levels 293

A.2 Requirements on video decoder capability 293

A.3 Profiles 293

A.4 Tiers and levels 308

Annex B (normative) Byte stream format 318

B.1 General 318

B.2 Byte stream NAL unit syntax and semantics 318

B.3 Byte stream NAL unit decoding process 319

B.4 Decoder byte-alignment recovery (informative) 320

Annex C (normative) Hypothetical reference decoder 321

C.1 General 321

C.2 Operation of coded picture buffer (CPB) 327

C.3 Operation of the decoded picture buffer (DPB) 333

C.4 Bitstream conformance 335

C.5 Decoder conformance 337

Annex D (normative) Supplemental enhancement information 341

D.1 General 341

D.2 SEI payload syntax 341

D.3 SEI payload semantics 359

Annex E (normative) Video usability information 443

E.1 General 443

E.2 VUI syntax 443

E.3 VUI semantics 446

Annex F (normative) Common specifications for multi-layer extensions 470

F.1 General 470

F.2 References 470

F.3 Definitions 470

F.4 Abbreviated terms 473

F.5 Conventions 473

F.6 Bitstream and picture formats, partitionings, scanning processes, and neighbouring relationships 474

F.7 Syntax and semantics 474

F.8 Decoding process 542

F.9 Parsing process 559

F.10 Specification of bitstream subsets 559

F.11 Profiles, tiers, and levels 561

F.12 Byte stream format 564

F.13 Hypothetical reference decoder 564

F.14 Supplemental enhancement information 589

F.15 Video usability information 617

Annex G (normative) Multiview high efficiency video coding 620

G.1 General 620

G.2 References 620

G.3 Terms and definitions 620

G.4 Abbreviations 620

G.5 Conventions 620

G.6 Bitstream and picture formats, partitionings, scanning processes, and neighbouring relationships 620

G.7 Syntax and semantics 620

G.8 Decoding processes 620

G.9 Parsing process 622

G.10 Specification of bitstream subsets 622

G.11 Profiles, tiers, and levels 622

G.12 Byte stream format 628

G.13 Hypothetical reference decoder 628

G.14 Supplemental enhancement information 628

G.15 Video usability information 643

Annex H (normative) Scalable high efficiency video coding 644
H.1 General 644
H.2 References 644
H.3 Definitions 644
H.4 Abbreviations 644
H.5 Conventions 644
H.6 Bitstream and picture formats, partitionings, scanning processes, and neighbouring relationships 644
H.7 Syntax and semantics 644
H.8 Decoding processes 644
H.9 Parsing process 662
H.10 Specification of bitstream subsets 662
H.11 Profiles, tiers, and levels 663
H.12 Byte stream format 672
H.13 Hypothetical reference decoder 672
H.14 Supplemental enhancement information 672
H.15 Video usability information 672
Annex I (normative) 3D high efficiency video coding 673
I.1 General 673
I.2 References 673
I.3 Definitions 673
I.4 Abbreviations 674
I.5 Conventions 674
I.6 Bitstream and picture formats, partitionings, scanning processes, and neighbouring relationships 674
I.7 Syntax and semantics 677
I.8 Decoding process 711
I.9 Parsing process 774
I.10 Specification of bitstream subsets 781
I.11 Profiles, tiers, and levels 781

I.12 Byte stream format 783

I.13 Hypothetical reference decoder 783

I.14 Supplemental enhancement information 783

Bibliography 795

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.

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 on 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 identical text is published as ITU-T H.265.

This third edition cancels and replaces the second edition (ISO/IEC 23008-2:2015), which has been technically revised. It also incorporates the Amendment ISO/IEC 23008-2:2015/Amd 1:2015.

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 fourth version of this document and the third 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 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, additional supplement enhancement information, 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 corresponds to the fourth 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 economic 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 start with Clause 1 (Scope) and move 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.

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*