This is a preview - click here to buy the full publication

INTERNATIONAL STANDARD

ISO/IEC 23008-2

Fifth edition
2023-10

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 23008-2:2023(E)

© ISO/IEC 2023

This is a preview - click here to buy the full publication

COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2023

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

Contents

Page

Foreword ..... **vii**
Introduction ..... **viii**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Abbreviated terms **22**
5 Conventions **24**
5.1 General 24
5.2 Arithmetic operators 24
5.3 Logical operators 25
5.4 Relational operators 25
5.5 Bit-wise operators 25
5.6 Assignment operators 26
5.7 Range notation 26
5.8 Mathematical functions 26
5.9 Order of operation precedence 27
5.10 Variables, syntax elements, and tables 28
5.11 Text description of logical operations 29
5.12 Processes 30
6 Bitstream and picture formats, partitionings, scanning processes, and neighbouring relationships **31**
6.1 Bitstream formats 31
6.2 Source, decoded, and output picture formats 31
6.3 Partitioning of pictures, slices, slice segments, tiles, CTUs, and CTBs 34
6.3.1 Partitioning of pictures into slices, slice segments, and tiles 34
6.3.2 Block and quadtree structures 36
6.3.3 Spatial or component-wise partitionings 37
6.4 Availability processes 37
6.4.1 Derivation process for z-scan order block availability 37
6.4.2 Derivation process for prediction block availability 38
6.5 Scanning processes 40
6.5.1 CTB raster and tile scanning conversion process 40
6.5.2 Z-scan order array initialization process 41
6.5.3 Up-right diagonal scan order array initialization process 41
6.5.4 Horizontal scan order array initialization process 42
6.5.5 Vertical scan order array initialization process 42

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

6.5.6 Traverse scan order array initialization process 43
7 Syntax and semantics 43
7.1 Method of specifying syntax in tabular form 43
7.2 Specification of syntax functions and descriptors 44
7.3 Syntax in tabular form 46
7.3.1 NAL unit syntax 46
7.3.2 Raw byte sequence payloads, trailing bits, and byte alignment syntax 47
7.3.3 Profile, tier and level syntax 56
7.3.4 Scaling list data syntax 59
7.3.5 Supplemental enhancement information message syntax 60
7.3.6 Slice segment header syntax 60
7.3.7 Short-term reference picture set syntax 65
7.3.8 Slice segment data syntax 66
7.4 Semantics 81
7.4.1 General 81
7.4.2 NAL unit semantics 81
7.4.3 Raw byte sequence payloads, trailing bits, and byte alignment semantics 92
7.4.4 Profile, tier, and level semantics 114
7.4.5 Scaling list data semantics 119
7.4.6 Supplemental enhancement information message semantics 121
7.4.7 Slice segment header semantics 122
7.4.8 Short-term reference picture set semantics 131
7.4.9 Slice segment data semantics 134
8 Decoding process 151
8.1 General decoding process 151
8.1.1 General 151
8.1.2 CVSG decoding process 151
8.1.3 Decoding process for a coded picture with nuh_layer_id equal to 0 152
8.2 NAL unit decoding process 154
8.3 Slice decoding process 154
8.3.1 Decoding process for picture order count 154
8.3.2 Decoding process for reference picture set 155
8.3.3 Decoding process for generating unavailable reference pictures 161
8.3.4 Decoding process for reference picture lists construction 162
8.3.5 Decoding process for collocated picture and no backward prediction flag 163
8.4 Decoding process for coding units coded in intra prediction mode 164
8.4.1 General decoding process for coding units coded in intra prediction mode 164

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

8.4.2 Derivation process for luma intra prediction mode 168
8.4.3 Derivation process for chroma intra prediction mode 171
8.4.4 Decoding process for intra blocks 171
8.5 Decoding process for coding units coded in inter prediction mode 184
8.5.1 General decoding process for coding units coded in inter prediction mode 184
8.5.2 Inter prediction process 185
8.5.3 Decoding process for prediction units in inter prediction mode 188
8.5.4 Decoding process for the residual signal of coding units coded in inter prediction mode 221
8.6 Scaling, transformation and array construction process prior to deblocking filter process 226
8.6.1 Derivation process for quantization parameters 226
8.6.2 Scaling and transformation process 228
8.6.3 Scaling process for transform coefficients 230
8.6.4 Transformation process for scaled transform coefficients 231
8.6.5 Residual modification process for blocks using a transform bypass 234
8.6.6 Residual modification process for transform blocks using cross-component prediction 235
8.6.7 Picture construction process prior to in-loop filter process 235
8.6.8 Residual modification process for blocks using adaptive colour transform 236
8.7 In-loop filter process 238
8.7.1 General 238
8.7.2 Deblocking filter process 239
8.7.3 Sample adaptive offset process 257
9 Parsing process 260
9.1 General 260
9.2 Parsing process for 0-th order Exp-Golomb codes 260
9.2.1 General 260
9.2.2 Mapping process for signed Exp-Golomb codes 262
9.3 CABAC parsing process for slice segment data 262
9.3.1 General 262
9.3.2 Initialization process 265
9.3.3 Binarization process 278
9.3.4 Decoding process flow 288
9.3.5 Arithmetic encoding process 304

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

10 Sub-bitstream extraction process 311
Annex A (normative) Profiles, tiers and levels 313
Annex B (normative) Byte stream format 343
Annex C (normative) Hypothetical reference decoder 346
Annex D (normative) Supplemental enhancement information 367
Annex E (normative) Video usability information 529
Annex F (normative) Common specifications for multi-layer extensions 559
Annex G (normative) Multiview high efficiency video coding 717
Annex H (normative) Scalable high efficiency video coding 743
Annex I (normative) 3D high efficiency video coding 774
Bibliography 898

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

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*, in collaboration with ITU-T (as Rec. ITU-T H.265).

This fifth edition cancels and replaces the fourth edition (ISO/IEC 23008-2:2020), which has been technically revised. It also incorporates the amendment ISO/IEC 23008-2:2020/Amd 1:2021.

The main changes are as follows:

— the specification of four additional “levels” of high capability for all profiles, referred to as levels 6.3, 7, 7.1, and 7.2;

— the specification of level 8.5 for the video profiles, which had previously been specified only for still picture profiles and which provides a way to identify bitstreams for which it is possible that the bitstreams do not conform to non-syntactical constraints specified for bitstream conformance to other levels.

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

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

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.

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

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.);

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

— 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 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 corresponded to the seventh version.

Rec. ITU-T H.265 | ISO/IEC 23008-2 version 8 refers to the integrated text additionally containing the shutter interval SEI message and corrections to various minor defects in the prior content of the specification.

Rec. ITU-T H.265 | ISO/IEC 23008-2 version 9 (the current version) refers to the integrated text additionally containing specification of levels 6.3, 7, 7.1, and 7.2, the specification of level 8.5 for the video profiles, and corrections to various minor defects in the prior content of the specification. This fifth edition published by ISO/IEC as ISO/IEC 23008-2:202x corresponds to the ninth version.

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

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

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

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 (Abbreviated terms), 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.

© ISO/IEC 2023 – All rights reserved

This is a preview - click here to buy the full publication

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)*

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

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

IETF RFC 5646, *Tags for Identifying Languages*

© ISO/IEC 2023 – All rights reserved