INTERNATIONAL STANDARD

ISO/IEC 14496-10

Ninth edition
2020-12

Information technology — Coding of audio-visual objects —

Part 10: Advanced video coding

Technologies de l'information — Codage des objets audiovisuels —

Partie 10: Codage visuel avancé

Reference number
ISO/IEC 14496-10:2020(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

Foreword vi

0 Introduction vii
0.1 Prologue vii
0.2 Purpose vii
0.3 Applications vii
0.4 Publication and versions of this document vii
0.5 Profiles and levels ix
0.6 Overview of the design characteristics ix
0.6.1 General ix
0.6.2 Predictive coding x
0.6.3 Coding of progressive and interlaced video x
0.6.4 Picture partitioning into macroblocks and smaller partitions x
0.6.5 Spatial redundancy reduction xi
0.7 How to read this document xi
0.8 Patent declarations xi

1 Scope 1

2 Normative references 1

3 Terms and definitions 1
3.1 General terms related to advanced video coding 1
3.2 Terms related to scalable video coding (Annex F) 16
3.3 Terms related to multiview video coding (Annex G) 22
3.4 Terms related to multiview and depth video coding (Annex H) 27
3.5 Terms related to multiview and depth video with enhanced non-base view coding (Annex I) 28

4 Abbreviated terms 28

5 Conventions 29
5.1 Arithmetic operators 29
5.2 Logical operators 30
5.3 Relational operators 30
5.4 Bit-wise operators 30
5.5 Assignment operators 31
5.6 Range notation 31
5.7 Mathematical functions 31
5.8 Order of operation precedence 32
5.9 Variables, syntax elements, and tables 33
5.10 Text description of logical operations 34
5.11 Processes 35

6 Source, coded, decoded and output data formats, scanning processes, and neighbouring relationships 35
6.1 Bitstream formats 35
6.2 Source, decoded, and output picture formats 35
6.3 Spatial subdivision of pictures and slices 40
6.4 Inverse scanning processes and derivation processes for neighbours 41
6.4.1 Inverse macroblock scanning process 41
6.4.2 Inverse macroblock partition and sub-macroblock partition scanning process 42
6.4.3 Inverse 4x4 luma block scanning process 43
6.4.4 Inverse 4x4 Cb or Cr block scanning process for ChromaArrayType equal to 3 44
6.4.5 Inverse 8x8 luma block scanning process 44
6.4.6 Inverse 8x8 Cb or Cr block scanning process for ChromaArrayType equal to 3 44
6.4.7 Inverse 4x4 chroma block scanning process 44
6.4.8 Derivation process of the availability for macroblock addresses 44
6.4.9 Derivation process for neighbouring macroblock addresses and their availability 45
6.4.10 Derivation process for neighbouring macroblock addresses and their availability in MBAFF frames 45
6.4.11 Derivation processes for neighbouring macroblocks, blocks, and partitions 46
6.4.12 Derivation process for neighbouring locations 51
6.4.13 Derivation processes for block and partition indices 54

7 Syntax and semantics 55
7.1 Method of specifying syntax in tabular form 55

7.2 Specification of syntax functions, categories, and descriptors 56
7.3 Syntax in tabular form 58
7.3.1 NAL unit syntax 58
7.3.2 Raw byte sequence payloads and RBSP trailing bits syntax 59
7.3.3 Slice header syntax 67
7.3.4 Slice data syntax 72
7.3.5 Macroblock layer syntax 73
7.4 Semantics 80
7.4.1 NAL unit semantics 80
7.4.2 Raw byte sequence payloads and RBSP trailing bits semantics 90
7.4.3 Slice header semantics 104
7.4.4 Slice data semantics 115
7.4.5 Macroblock layer semantics 116
8 Decoding process 129
8.1 NAL unit decoding process 130
8.2 Slice decoding process 131
8.2.1 Decoding process for picture order count 131
8.2.2 Decoding process for macroblock to slice group map 135
8.2.3 Decoding process for slice data partitions 138
8.2.4 Decoding process for reference picture lists construction 139
8.2.5 Decoded reference picture marking process 146
8.3 Intra prediction process 150
8.3.1 Intra_4x4 prediction process for luma samples 150
8.3.2 Intra_8x8 prediction process for luma samples 157
8.3.3 Intra_16x16 prediction process for luma samples 164
8.3.4 Intra prediction process for chroma samples 167
8.3.5 Sample construction process for I_PCM macroblocks 171
8.4 Inter prediction process 172
8.4.1 Derivation process for motion vector components and reference indices 174
8.4.2 Decoding process for Inter prediction samples 186
8.4.3 Derivation process for prediction weights 195
8.5 Transform coefficient decoding process and picture construction process prior to deblocking filter process 197
8.5.1 Specification of transform decoding process for 4x4 luma residual blocks 197
8.5.2 Specification of transform decoding process for luma samples of Intra_16x16 macroblock prediction mode 198
8.5.3 Specification of transform decoding process for 8x8 luma residual blocks 199
8.5.4 Specification of transform decoding process for chroma samples 200
8.5.5 Specification of transform decoding process for chroma samples with ChromaArrayType equal to 3 201
8.5.6 Inverse scanning process for 4x4 transform coefficients and scaling lists 202
8.5.7 Inverse scanning process for 8x8 transform coefficients and scaling lists 203
8.5.8 Derivation process for chroma quantization parameters 204
8.5.9 Derivation process for scaling functions 205
8.5.10 Scaling and transformation process for DC transform coefficients for Intra_16x16 macroblock type 206
8.5.11 Scaling and transformation process for chroma DC transform coefficients 207
8.5.12 Scaling and transformation process for residual 4x4 blocks 209
8.5.13 Scaling and transformation process for residual 8x8 blocks 211
8.5.14 Picture construction process prior to deblocking filter process 215
8.5.15 Intra residual transform-bypass decoding process 217
8.6 Decoding process for P macroblocks in SP slices or SI macroblocks 217
8.6.1 SP decoding process for non-switching pictures 218
8.6.2 SP and SI slice decoding process for switching pictures 220
8.7 Deblocking filter process 222
8.7.1 Filtering process for block edges 226
8.7.2 Filtering process for a set of samples across a horizontal or vertical block edge 228
9 Parsing process 234
9.1 Parsing process for Exp-Golomb codes 234
9.1.1 Mapping process for signed Exp-Golomb codes 236
9.1.2 Mapping process for coded block pattern 236
9.2 CAVLC parsing process for transform coefficient levels 239
9.2.1 Parsing process for total number of non-zero transform coefficient levels and number of trailing ones 240

9.2.2 Parsing process for level information 243
9.2.3 Parsing process for run information 245
9.2.4 Combining level and run information 248
9.3 CABAC parsing process for slice data 248
9.3.1 Initialization process 249
9.3.2 Binarization process 273
9.3.4 Arithmetic encoding process 303
Annex A (normative) Profiles and levels **310**
Annex B (normative) Byte stream format **333**
Annex C (normative) Hypothetical reference decoder **336**
Annex D (normative) Supplemental enhancement information **357**
Annex E (normative) Video usability information **445**
Annex F (normative) Scalable video coding **465**
Annex G (normative) Multiview video coding **689**
Annex H (normative) Multiview and depth video coding **755**
Annex I (normative) Multiview and depth video with enhanced non-base view coding **804**
Bibliography **859**

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*, in collaboration with ITU-T. The technically identical text is published as ITU-T H.264 (06/2019).

This ninth edition cancels and replaces the eighth edition (ISO/IEC 14496-10:2014), which has been technically revised. It also incorporates the Amendments ISO/IEC 14496-10:2014/Amd. 1:2015 and ISO/IEC 14496-10:2014/Amd. 3:2016.

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

— specification of an additional profile (the Progressive High 10 profile);

— additional colour-related video usability information codepoint identifiers;

— additional supplemental enhancement information messages;

— minor corrections and clarifications throughout the document.

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.

0 Introduction

0.1 Prologue

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 Video Team (JVT) in 2001 for development of a new Recommendation | International Standard.

0.2 Purpose

This Recommendation | International Standard 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 communication. It is also designed to enable the use of the coded video representation in a flexible manner for a wide variety of network environments. The use of this Recommendation | International Standard 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.

0.3 Applications

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

- CATV: cable TV on optical networks, copper, etc.
- DBS: direct broadcast satellite video services.
- DSL: digital subscriber line video services.
- DTTB: digital terrestrial television broadcasting.
- ISM: interactive storage media (optical disks, etc.).
- MMM: multimedia mailing.
- MSPN: multimedia services over packet networks.
- RTC: real-time conversational services (videoconferencing, videophone, etc.).
- RVS: remote video surveillance.
- SSM: serial storage media (digital VTR, etc.).

0.4 Publication and versions of this document

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 1 refers to the first approved version of this Recommendation | International Standard.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 2 refers to the integrated text containing the corrections specified in the first technical corrigendum.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 3 refers to the integrated text containing both the first technical corrigendum (2004) and the first amendment, which is referred to as the "Fidelity range extensions".

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 4 refers to the integrated text containing the first technical corrigendum (2004), the first amendment (the "Fidelity range extensions"), and an additional technical corrigendum (2005).

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 5 refers to the integrated version 4 text with its specification of the High 4:4:4 profile removed.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 6 refers to the integrated version 5 text after its amendment to support additional colour space indicators.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 7 refers to the integrated version 6 text after its amendment to define five new profiles intended primarily for professional applications (the High 10 Intra, High 4:2:2 Intra, High 4:4:4 Intra, CAVLC 4:4:4 Intra, and High 4:4:4 Predictive profiles) and two new types of supplemental enhancement information (SEI) messages (the post-filter hint SEI message and the tone mapping information SEI message).

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 8 refers to the integrated version 7 text after its amendment to specify scalable video coding in three profiles (Scalable Baseline, Scalable High, and Scalable High Intra profiles).

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 9 refers to the integrated version 8 text after applying the corrections specified in a third technical corrigendum.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 10 refers to the integrated version 9 text after its amendment to specify a profile for multiview video coding (the Multiview High profile) and to define additional SEI messages.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 11 refers to the integrated version 10 text after its amendment to define a new profile (the Constrained Baseline profile) intended primarily to enable implementation of decoders supporting only the common subset of capabilities supported in various previously-specified profiles.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 12 refers to the integrated version 11 text after its amendment to define a new profile (the Stereo High profile) for two-view video coding with support of interlaced coding tools and to specify an additional SEI message specified as the frame packing arrangement SEI message. The changes for versions 11 and 12 were processed as a single amendment in the ISO/IEC approval process.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 13 refers to the integrated version 12 text with various minor corrections and clarifications as specified in a fourth technical corrigendum.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 14 refers to the integrated version 13 text after its amendment to define a new level (Level 5.2) supporting higher processing rates in terms of maximum macroblocks per second and a new profile (the Progressive High profile) to enable implementation of decoders supporting only the frame coding tools of the previously-specified High profile.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 15 refers to the integrated version 14 text with miscellaneous corrections and clarifications as specified in a fifth technical corrigendum.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 16 refers to the integrated version 15 text after its amendment to define three new profiles intended primarily for communication applications (the Constrained High, Scalable Constrained Baseline, and Scalable Constrained High profiles).

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 17 refers to the integrated version 16 text after its amendment to define additional supplemental enhancement information (SEI) message data, including the multiview view position SEI message, the display orientation SEI message, and two additional frame packing arrangement type indication values for the frame packing arrangement SEI message (the 2D content and tiled arrangement type indication values).

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 18 refers to the integrated version 17 text after its amendment to specify the coding of depth signals, including the specification of an additional profile, the Multiview Depth High profile.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 19 refers to the integrated version 18 text after incorporating a correction to the sub-bitstream extraction process for multiview video coding.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 20 refers to the integrated version 19 text after its amendment to specify the combined coding of video view and depth enhancement, including the specification of an additional profile, the Enhanced Multiview Depth High profile.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 21 refers to the integrated version 20 text after its amendment to specify additional colorimetry identifiers and an additional model type in the tone mapping information SEI message.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 22 refers to the integrated version 21 text after its amendment to specify multi-resolution frame-compatible (MFC) enhancement for stereoscopic video coding, including the specification of an additional profile, the MFC High profile.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 23 refers to the integrated version 22 text after its amendment to specify multi-resolution frame-compatible (MFC) stereoscopic video with depth maps, including the specification of an additional profile, the MFC Depth High profile, and the mastering display colour volume SEI message, additional colour-related video usability information codepoint identifiers, and miscellaneous minor corrections and clarifications.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 24 refers to the integrated version 23 text after its amendment to specify additional levels of decoder capability supporting larger picture sizes (Levels 6, 6.1, and 6.2), the green metadata SEI message, the alternative depth information SEI message, additional colour-related video usability information codepoint identifiers, and miscellaneous minor corrections and clarifications.

Rec. ITU-T H.264 | ISO/IEC 14496-10 version 25 refers to the integrated version 24 text after its amendment to specify the Progressive High 10 profile; support for additional colour-related indicators, including the hybrid log-gamma transfer characteristics indication, the alternative transfer characteristics SEI message, the IC_TC_P colour matrix transformation,

chromaticity-derived constant luminance and non-constant luminance colour matrix coefficients, the colour remapping information SEI message, and miscellaneous minor corrections and clarifications.

Rec. ITU-T H.264 | ISO/IEC 14496-10 version 26 (the current document) refers to the integrated version 25 text after its amendment to specify additional SEI messages for ambient viewing environment, content light level information, content colour volume, equirectangular projection, cubemap projection, sphere rotation, region-wise packing, omnidirectional viewport, SEI manifest, and SEI prefix, and miscellaneous minor corrections and clarifications.

This edition corresponds in technical content to the thirteenth edition in ITU-T (approved in June 2019).

0.5 Profiles and levels

This Recommendation | International Standard 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" 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 by this Recommendation | International Standard. 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, "levels" are specified within each profile. A level 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).

Coded video content conforming to this Recommendation | International Standard 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.

0.6 Overview of the design characteristics

0.6.1 General

The coded representation specified in the syntax is designed to enable a high compression capability for a desired image quality. With the exception of the transform bypass mode of operation for lossless coding in the High 4:4:4 Intra, CAVLC 4:4:4 Intra, and High 4:4:4 Predictive profiles, and the I_PCM mode of operation in all profiles, 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 is then further compressed using a transform to remove spatial correlation inside the transform block before it is quantized, producing an 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 are combined with the quantized transform coefficient information and encoded using either variable length coding or arithmetic coding.

Scalable video coding is specified in Annex F allowing the construction of bitstreams that contain sub-bitstreams that conform to this document. For temporal bitstream scalability, i.e., the presence of a sub-bitstream with a smaller temporal sampling rate than the bitstream, complete access units are removed from the bitstream when deriving the sub-bitstream. In this case, high-level syntax and inter prediction reference pictures in the bitstream are constructed accordingly. For spatial and quality bitstream scalability, i.e., the presence of a sub-bitstream with lower spatial resolution or quality than the bitstream, NAL units are removed from the bitstream when deriving the sub-bitstream. In this case, inter-layer prediction, i.e., the prediction of the higher spatial resolution or quality signal by data of the lower spatial resolution or

quality signal, is typically used for efficient coding. Otherwise, the coding algorithm as described in the previous paragraph is used.

Multiview video coding is specified in Annex G allowing the construction of bitstreams that represent multiple views. Similar to scalable video coding, bitstreams that represent multiple views may also contain sub-bitstreams that conform to this document. For temporal bitstream scalability, i.e., the presence of a sub-bitstream with a smaller temporal sampling rate than the bitstream, complete access units are removed from the bitstream when deriving the sub-bitstream. In this case, high-level syntax and inter prediction reference pictures in the bitstream are constructed accordingly. For view bitstream scalability, i.e. the presence of a sub-bitstream with fewer views than the bitstream, NAL units are removed from the bitstream when deriving the sub-bitstream. In this case, inter-view prediction, i.e., the prediction of one view signal by data of another view signal, is typically used for efficient coding. Otherwise, the coding algorithm as described in the previous paragraph is used.

An extension of multiview video coding that additionally supports the inclusion of depth maps is specified in Annex H, allowing the construction of bitstreams that represent multiple views with corresponding depth views. In a similar manner as with the multiview video coding specified in Annex G, bitstreams encoded as specified in Annex H may also contain sub-bitstreams that conform to this document.

A multiview video coding extension with depth information is specified in Annex I. Sub-bitstreams consisting of a texture base view conform to this document, sub-bitstreams consisting of multiple texture views may also conform to Annex G of this document, and sub-bitstreams consisting of one or more texture views and one or more depth views may also conform to Annex H of this document. Enhanced texture view coding that utilizes the associated depth views and decoding processes for depth views are specified for this extension.

0.6.2 Predictive coding

Because of the conflicting requirements of random access and highly efficient compression, two main coding types are specified. Intra coding is done without reference to other pictures. Intra coding may provide access points to the coded sequence where decoding can begin and continue correctly, but typically also shows only moderate compression efficiency. Inter coding (predictive or bi-predictive) is more efficient using inter prediction of each block of sample values from some previously decoded picture selected by the encoder. In contrast to some other video coding standards, pictures coded using bi-predictive inter prediction may also be used as references for inter coding of other pictures.

The application of the three coding types to pictures in a sequence is flexible, and the order of the decoding process is generally not the same as the order of the source picture capture process in the encoder or the output order from the decoder for display. The choice is left to the encoder and will depend on the requirements of the application. The decoding order is specified such that the decoding of pictures that use inter-picture prediction follows later in decoding order than other pictures that are referenced in the decoding process.

0.6.3 Coding of progressive and interlaced video

This Recommendation | International Standard specifies a syntax and decoding process for video that originated in either progressive-scan or interlaced-scan form, which may be mixed together in the same sequence. The two fields of an interlaced frame are separated in capture time while the two fields of a progressive frame share the same capture time. Each field may be coded separately or the two fields may be coded together as a frame. Progressive frames are typically coded as a frame. For interlaced video, the encoder can choose between frame coding and field coding. Frame coding or field coding can be adaptively selected on a picture-by-picture basis and also on a more localized basis within a coded frame. Frame coding is typically preferred when the video scene contains significant detail with limited motion. Field coding typically works better when there is fast picture-to-picture motion.

0.6.4 Picture partitioning into macroblocks and smaller partitions

As in previous video coding Recommendations and International Standards, a macroblock, consisting of a 16x16 block of luma samples and two corresponding blocks of chroma samples, is used as the basic processing unit of the video decoding process.

A macroblock can be further partitioned for inter prediction. The selection of the size of inter prediction partitions is a result of a trade-off between the coding gain provided by using motion compensation with smaller blocks and the quantity of data needed to represent the data for motion compensation. In this Recommendation | International Standard the inter prediction process can form segmentations for motion representation as small as 4x4 luma samples in size, using motion vector accuracy of one-quarter of the luma sample grid spacing displacement. The process for inter prediction of a sample block can also involve the selection of the picture to be used as the reference picture from a number of stored previously-decoded pictures. Motion vectors are encoded differentially with respect to predicted values formed from nearby encoded motion vectors.

Typically, the encoder calculates appropriate motion vectors and other data elements represented in the video data stream. This motion estimation process in the encoder and the selection of whether to use inter prediction for the representation of each region of the video content is not specified in this Recommendation | International Standard.

0.6.5 Spatial redundancy reduction

Both source pictures and prediction residuals have high spatial redundancy. This Recommendation | International Standard is based on the use of a block-based transform method for spatial redundancy removal. After inter prediction from previously-decoded samples in other pictures or spatial-based prediction from previously-decoded samples within the current picture, the resulting prediction residual is split into 4x4 blocks. These are converted into the transform domain where they are quantized. After quantization many of the transform coefficients are zero or have low amplitude and can thus be represented with a small amount of encoded data. The processes of transformation and quantization in the encoder are not specified in this Recommendation | International Standard.

0.7 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 subclauses 7.1-7.3 for syntactical order and see subclause 7.4 for semantics; i.e., 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). 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 G also form an integral part of this Recommendation | International Standard.

Annex A specifies fourteen profiles (Baseline, Constrained Baseline, Main, Extended, High, Progressive High, Constrained High, High 10, High 4:2:2, High 4:4:4 Predictive, High 10 Intra, High 4:2:2 Intra, High 4:4:4 Intra, and CAVLC 4:4:4 Intra), each being tailored to certain application domains, and defines the so-called 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 and its use 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 scalable video coding (SVC). The reader is referred to Annex F for the entire decoding process for SVC, which is specified there with references being made to Clauses 2-9 and Annexes A-E. Subclause F.10 specifies five profiles for SVC (Scalable Baseline, Scalable Constrained Baseline, Scalable High, Scalable Constrained High, and Scalable High Intra).

Annex G specifies multiview video coding (MVC) and multi-resolution frame compatible stereo coding (MFC). The reader is referred to Annex G for the entire decoding process for MVC and MFC, which is specified there with references being made to Clauses 2-9 and Annexes A-E. Subclause G.10 specifies two profiles for MVC (Multiview High and Stereo High) and one profile for MFC (MFC High).

Annex H specifies MVC extensions for inclusion of depth maps, referred to as multiview video coding with depth (MVCD). The reader is referred to Annex H for the entire decoding process for MVCD, which is specified there with references being made to Clauses 2-9 and Annexes A-E and Annex G. Subclause H.10 specifies two profiles for MVCD (Multiview Depth High and MFC Depth High).

Annex I specifies a multiview video coding extension with depth information (3D-AVC). The reader is referred to Annex I for the entire decoding process for 3D-AVC, which is specified there with references being made to Clauses 2-9 and Annexes A-E and Annexes G-H. Subclause I.10 specifies one profile for 3D-AVC.

0.8 Patent declarations

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 — Coding of audio-visual objects —

Part 10:
Advanced video coding

1 Scope

This document specifies advanced video coding for coding of audio-visual objects.

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.

ITU-T Recommendation T.35 (2000), *Procedure for the allocation of ITU-T defined codes for non-standard facilities*

ISO/IEC 11578:1996, *Information technology — Open Systems Interconnection — Remote Procedure Call (RPC)*

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

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

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

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

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

3.1 General terms related to advanced video coding

3.1.1
access unit
set of *NAL units* that are consecutive in *decoding order* and contain exactly one *primary coded picture*

Note 1 to entry: In addition to the primary coded picture, an access unit may also contain one or more redundant coded pictures, one auxiliary coded picture, or other NAL units not containing slices or slice data partitions of a coded picture. The decoding of an access unit always results in a decoded picture.

3.1.2
AC transform coefficient
any *transform coefficient* for which the *frequency index* in one or both dimensions is non-zero

3.1.3
adaptive binary arithmetic decoding process
entropy *decoding process* that derives the values of *bins* from a *bitstream* produced by an *adaptive binary arithmetic encoding process*

3.1.4
adaptive binary arithmetic encoding process
entropy *encoding process* that codes a sequence of *bins* and produces a *bitstream* that can be decoded using the *adaptive binary arithmetic decoding process*

Note 1 to entry: The encoding process is not specified in this Recommendation | International Standard.

3.1.5
alpha blending
process in which an *auxiliary coded picture* is used in combination with a *primary coded picture* and with other data in the *display process*

Note 1 to entry: In an alpha blending process, the samples of an auxiliary coded picture are interpreted as indications of the degree of opacity (or, equivalently, the degrees of transparency) associated with the corresponding luma samples of the primary coded picture.

Note 2 to entry: The alpha blending process and other data are not specified by this Recommendation | International Standard.

3.1.6
arbitrary slice order
ASO
decoding order of *slices* in which the *macroblock address* of the first *macroblock* of some *slice* of a *slice group* may be less than the *macroblock address* of the first *macroblock* of some other preceding *slice* of the same *slice group* or, in the case of a *picture* that is coded using three separate colour planes, some other preceding *slice* of the same *slice group* within the same colour plane, or in which the *slices* of a *slice group* of a picture may be interleaved with the *slices* of one or more other *slice groups* of the *picture* or, in the case of a *picture* that is coded using three separate colour planes, with the *slices* of one or more other *slice groups* within the same colour plane

3.1.7
auxiliary coded picture
picture that supplements the *primary coded picture* that may be used in combination with other data in the *display process*

Note 1 to entry: An auxiliary coded picture has the same syntactic and semantic restrictions as a monochrome redundant coded picture. An auxiliary coded picture must contain the same number of macroblocks as the primary coded picture. Auxiliary coded pictures have no normative effect on the decoding process. See also primary coded picture and redundant coded picture.

Note 2 to entry: The other data is not specified by this Recommendation | International Standard.

3.1.8
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.

3.1.9
B slice
bi-predictive slice
slice that may be decoded using *intra prediction* or *inter prediction* using at most two *motion vectors* and *reference indices* to *predict* the sample values of each *block*

3.1.10
bin
one bit of a *bin string*

3.1.11
binarization
set of *bin strings* for all possible values of a *syntax element*

3.1.12
binarization process
unique mapping process of all possible values of a *syntax element* onto a set of *bin strings*

3.1.13
bin string
string of bins
intermediate binary representation of values of *syntax elements* from the *binarization* of the *syntax element*

3.1.14
bitstream
sequence of bits that forms the representation of *coded pictures* and associated data forming one or more *coded video sequences*

Note 1 to entry: Bitstream is a collective term used to refer either to a NAL unit stream or a byte stream.

3.1.15
block
MxN (M-column by N-row) array of samples, or an MxN array of *transform coefficients*

3.1.16
bottom field
one of two *fields* that comprise a *frame*

Note 1 to entry: Each row of a bottom field is spatially located immediately below a corresponding row of a top field.

3.1.17
bottom macroblock
macroblock within a *macroblock pair* that contains the samples in the bottom row of samples for the *macroblock pair*

Note 1 to entry: For a field macroblock pair, the bottom macroblock represents the samples from the region of the bottom field of the frame that lie within the spatial region of the macroblock pair. For a frame macroblock pair, the bottom macroblock represents the samples of the frame that lie within the bottom half of the spatial region of the macroblock pair.

3.1.18
broken link
location in a *bitstream* at which it is indicated that some subsequent *pictures* in *decoding order* may contain serious visual artefacts due to unspecified operations performed in the generation of the *bitstream*

3.1.19
byte
sequence of 8 bits, written and read with the most significant bit on the left and the least significant bit on the right

Note 1 to entry: When represented in a sequence of data bits, the most significant bit of a byte is first.

3.1.20
byte-aligned
position in a *bitstream* is byte-aligned when the position is an integer multiple of 8 bits from the position of the first bit in the *bitstream*

Note 1 to entry: A bit or byte or syntax element is said to be byte-aligned when the position at which it appears in a bitstream is byte-aligned.

3.1.21
byte stream
encapsulation of a *NAL unit stream* containing *start code prefixes* and *NAL units*

Note 1 to entry: As specified in Annex B.

3.1.22
category
number associated with each *syntax element*

Note 1 to entry: The category is used to specify the allocation of syntax elements to NAL units for slice data partitioning. It may also be used in a manner determined by the application to refer to classes of syntax elements in a manner not specified in this Recommendation | International Standard.