INTERNATIONAL STANDARD

ISO/IEC 14496-33

First edition
2019-02

Information technology — Coding of audio-visual objects —

Part 33: Internet video coding

Technologies de l'information — Codage des objets audiovisuels — Partie 33: Codage vidéo Internet

Reference number
ISO/IEC 14496-33:2019(E)

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

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. 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 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 Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

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

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document specifies Internet video coding, a video compression technology that is intended to be suitable for video distribution models currently adopted on the Internet.

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:

Nokia Technologies Oy

Joensuunkatu 7E

FIN-24100 Salo

FINLAND

Telephone : +358 50 366 2022

Apple Inc.

Intellectual Property and Licensing

1 Infinite Loop, MS 169-3IPL

Cupertino, CA 95014

USA

Telephone: +1(408) 974-0015

Industry-University Cooperation Foundation Hanyang University

222 Wangsimni-ro, Seongdong-gu

Seoul 04763

REPUBLIC OF KOREA

Telephone: +82-2-2220-2212

Mitsubishi Electric Corporation
Corporate Licensing Division
2-7-3 Marunouchi, Chiyoda-ku
Tokyo 100-8310
JAPAN
Telephone: +81-3-3218-3465

QUALCOMM Incorporated
5775 Morehouse Drive
San Diego, CA 92121
USA
Telephone: +1 (858) 587-1121

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

Information technology — Coding of audio-visual objects —

Part 33:
Internet video coding

1 Scope

This document specifies MPEG-4 Internet 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.

Rec. ITU-T H.262 | ISO/IEC 13818-2: 2013, *Information technology — Generic coding of moving pictures and associated audio information — Part 2: Video*

IEC 60461, *Time and control code*

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:

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

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

3.1
B frame
bidirectional frame
frame (3.28) that is coded using motion compensated prediction from past or future *reference frames* (3.53) in *output order* (3.40)

3.2
backward prediction
process of predicting the current *frame* (3.28) by using future *frames* in an *output order* (3.40) as *reference frames* (3.53)

3.3
bin
bit of a *bin string* (3.4)

3.4
bin string
intermediate binary representation of values of *syntax elements* (3.65) resulting from the *binarization* (3.5) of the *syntax element*

3.5
binarization
set of *bin strings* (3.4) for all possible values of a *syntax element* (3.65)

3.6
binarization process
unique mapping process of all possible values of a *syntax element* (3.65) onto a set of *bin strings* (3.4)

3.7
bitstream
ordered series of bits that forms the *coded representation* (3.14) of the data

3.8
block
MxN (M-column by N-row) array of samples, or an MxN array of *transform coefficients* (3.66)

3.9
byte
sequence of 8 bits, written and read with the most significant bit on the left and the least significant bit on the right, such that when represented in a sequence of data bits, the most significant bit of a byte is first

3.10
byte-aligned
positioning of a bit or *byte* (3.9) or *syntax element* (3.65) when the position at which it appears in a *bitstream* (3.7) is an integer multiple of 8 bits from the position of the first bit in the *bitstream*

3.11
byte stream
ordered series of bytes that forms the *coded representation* (3.14) of the data

3.12
chroma
sample array or single sample, identified symbolically by Cb or Cr, representing one of the two colour difference signals related to the primary colours

Note 1 to entry: The term chroma is used rather than the term chrominance in order to avoid the implication of the use of linear light transfer characteristics that is often associated with the term chrominance.

3.13
coded frame
coded representation (3.14) of a *frame* (3.28)

3.14
coded representation
series of data elements as represented in coded form in the *bitstream* (3.7)

3.15
component
array or single sample from one of the three arrays (*luma* (3.37) and two *chroma* (3.12)) that make up a *frame* (3.28) in 4:2:0 colour format

3.16
DC coefficient
transform coefficient (3.66) for which the *frequency index* (3.27) is zero in all dimensions

3.17
decoded frame
frame (3.28) derived by decoding a *coded frame* (3.13)

3.18
decoder
embodiment of the *decoding process* (3.20)

3.19
decoding order
order in which syntax elements are processed by the *decoding process* (3.20)

3.20
decoding process
process that derives *decoded frames* (3.17) from the syntax elements in the *bitstream* (3.7)

3.21
dequantization
process of *scaling* (3.57) the *quantized transform coefficients* (3.49) after their representation in the *bitstream* (3.7) has been *parsed* (3.42) and before they are presented to the *inverse transform* (3.34) part of the *decoding process* (3.20)

3.22
encoder
embodiment of an *encoding process* (3.23)

3.23
encoding process
process that produces a *bitstream* (3.7)

Note 1 to entry: This document does not specify an encoding process.

3.24
forbidden
specification that a value shall never be used

Note 1 to entry: This is usually to avoid emulation of a *start code* (3.63) pattern.

3.25
forward prediction
process of predicting the current frame by the past *reference frames* (3.53) in output order

3.26
flag
binary variable that can take one of the two possible values, 0 and 1

3.27
frequency index
one-dimensional or two-dimensional index associated with a *transform coefficient* (3.66) prior to an *inverse transform* (3.34) part of a *decoding process* (3.20)

3.28
frame
successive lines, numbered from the top-most line to the bottom-most line, containing samples numbered from the left-most sample to the right-most sample, representing the spatial information of a video signal from a single time instant

3.29
I frame
intra frame
frame (3.28) coded using information only from itself

3.30
inter macroblock
macroblock (3.38) which is coded using *inter prediction* (3.31)

3.31
inter prediction
prediction (3.44) derived from data elements (e.g. sample value or *motion vector* (3.39)) of *reference frames* (3.53) other than the current frame

3.32
intra macroblock
macroblock (3.38) which is coded using *intra prediction* (3.33)

3.33
intra prediction
prediction (3.44) derived from only data elements (e.g. sample values) of the same decoded *slice* (3.60)

3.34
inverse transform
part of the *decoding process* (3.20) by which a set of *transform coefficients* (3.66) are converted into spatial-domain values, or by which a set of *transform coefficients* are converted into *DC coefficients* (3.16)

3.35
layer
one of a set of syntactical structures in a non-branching hierarchical relationship, such that higher layers contain lower layers, with such coded layers being the *coded frame* (3.13), *slice* (3.60), *macroblock* (3.38) and *block* (3.8)

3.36
level
defined set of constraints on the values that may be taken by *syntax elements* (3.65) and variables; or in a different context, the value of a *transform coefficient* (3.66) prior to *scaling* (3.57)

Note 1 to entry: The same set of levels is defined for all *profiles* (3.47), with most aspects of the definition of each level being in common across different *profiles*. Individual implementations may, within specified constraints, support a different level for each supported *profile*.

3.37
luma
sample array or single sample, identified symbolically by Y or L, ordinarily representing the brightness signal related to the primary colours

Note 1 to entry: The term luma is used rather than the term luminance in order to avoid the implication of the use of linear light transfer characteristics that is often associated with the term luminance. The symbol L is sometimes used instead of the symbol Y to avoid confusion with the symbol y as used for vertical location.

3.38
macroblock
16 × 16 *luma* (3.37) sample value block and its corresponding two *chroma* (3.12) sample value blocks

3.39
motion vector
two-dimensional vector used for *inter prediction* (3.31) that provides an offset from the coordinates in the *decoded frame* (3.17) to the coordinates in a *reference frame* (3.53)

3.40
output order
order in which the *decoded frames* (3.17) are output from the *decoded frame* buffer in case the *decoded frames* are to be output from the *decoded frame* buffer

3.41
P frame
predictive frame
frame (3.28) that is coded using motion compensated prediction from past *reference frames* (3.53) in *output order* (3.40)

3.42
parse
procedure of obtaining the value of a *syntax element* (3.65) from a *bitstream* (3.7)

3.43
partitioning
division of a set into subsets such that each element of a set is in exactly one of the subsets

3.44
prediction
embodiment of a *prediction process* (3.45)

3.45
prediction process
use of a *predictor* (3.46) to provide an estimate of a data element (e.g. sample value or *motion vector* (3.39)) currently being decoded

3.46
predictor
combination of specified values or previously decoded data elements (e.g. sample value or *motion vector* (3.39)) used in the *decoding process* (3.20) of subsequent data elements

3.47
profile
specified subset of the syntax

3.48
quantization parameter
variable used by the *decoding process* (3.20) for *scaling* (3.57) of *transform coefficient levels* (3.67)

3.49
quantized transform coefficients
transform coefficients (3.66) before *dequantization* (3.21)

3.50
random access
starting the *decoding process* (3.20) for part of a *bitstream* at some point other than the beginning of the *bitstream* (3.7)

3.51
raster scan
mapping of a rectangular two-dimensional pattern to a one-dimensional pattern such that the first entries in the one-dimensional pattern are from the top-most row of the two-dimensional pattern scanned from left to right, followed similarly by the second, third, etc., top-most rows of the pattern (proceeding downwards), with each row scanned from left to right

3.52
reference index
order indication of the *reference frames* (3.53) in the frame buffer in the *decoding process* (3.20)

3.53
reference frame
frame that contains samples that may be used for *inter prediction* (3.31) in the *decoding process* (3.20) of subsequent *frames* (3.28) in *decoding order* (3.19)

3.54
reserved
specification that some values of a particular *syntax element* (3.65) are for future use by ISO/IEC, such that these values shall not be used in *bitstreams* (3.7), but may be specified for use in future extensions by ISO/IEC

3.55
residual
decoded difference between a *prediction* (3.44) of a sample or data element and its decoded value

3.56
run
number of data elements with the same value or the same treatment in the *decoding process* (3.20)

Note 1 to entry: In one context, it means the number of zero coefficients before a non-zero coefficient in the block scan, and in another context, it means the number of consecutive *skipped macroblocks* (3.59).

3.57
scaling
process of multiplying *transform coefficient levels* (3.67) by a factor, resulting in *transform coefficients* (3.66)

3.58
sequence
highest layer syntax structure of the *bitstream* (3.7), including one or more consecutive *coded frames* (3.13)

3.59
skipped macroblock
macroblock (3.38) for which no syntax elements are present in the *bitstream* (3.7) except for the indication that the *macroblock* is a *skipped macroblock* (3.59)

3.60
slice
integer number of consecutive *macroblock* (3.38) rows in the *raster scan* (3.51) order that is associated with the same header data

3.61
slice header
part of a coded *slice* (3.60) containing the data elements pertaining to the first or all *macroblocks* (3.58) represented in a *slice*

3.62
source
video material or some of its attributes before operation of an *encoding process* (3.23)

3.63
start code
32-bit codeword pattern which is unique in the whole *bitstream* (3.7)

Note 1 to entry: Start codes can be used to identify the starting point of a syntax structure in the *bitstream* (e.g. to enable *random access* (3.50)).

3.64
stuffing bits
bit string having a prescribed pattern of fixed values at a particular position in the *bitstream* (3.7)

3.65
syntax element
element of data represented in the *bitstream* (3.7)

3.66
transform coefficient
scalar quantity, considered to be in a frequency domain, that is associated with a particular one-dimensional or two-dimensional *frequency index* (3.27) in an *inverse transform* (3.34) part of the *decoding process* (3.20)

