INTERNATIONAL STANDARD

ISO/IEC 21122-2

First edition
2019-07

Information technology — JPEG XS low-latency lightweight image coding system —

Part 2:
Profiles and buffer models

COPYRIGHT PROTECTED DOCUMENT

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) 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*.

A list of all parts in the ISO/IEC 21122 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

ISO/IEC 21122-1 (JPEG XS) specifies a single syntax designed to serve a wide range of applications, bit rates, resolutions, qualities, and services. Its main target applications are video transport over video links and IP networks, real-time video storage, video memory buffer, omni-directional video capture system, head-mounted displays for virtual or augmented reality and sensor compression for the automotive industry. These applications have different requirements in terms of complexity, latency and compression efficiency. Even within a given application field, different requirements are usually identified depending on the targeted use case.

Considering the impracticality of implementing the full syntax of ISO/IEC 21122-1, and in order to meet the requirements of the different target applications while safeguarding as much as possible the interoperability enabled by the common syntax defined in ISO/IEC 21122-1, a limited number of subsets of this syntax are stipulated by means of "profiles", "levels", and "sublevels".

The coding tools specified in ISO/IEC 21122-1 allow encoder and decoder implementations to limit the end-to-end latency to a fraction of the frame size. To ensure this property, this document specifies a buffer model, consisting of a decoder model and a transmission channel model.

iTeh Standards
(https://standards.iteh.ai)
Document Preview

ISO/IEC 21122-2:2019
https://standards.iteh.ai/catalog/standards/iso/a8d09a19-f7b3-46bc-a9aa-eed70c73325e/iso-iec-21122-2-2019

Information technology — JPEG XS low-latency lightweight image coding system —

Part 2:
Profiles and buffer models

1   Scope

This document defines a limited number of subsets of the syntax specified in ISO/IEC 21122-1 and a buffer model to ensure interoperability between implementations in the presence of a latency constraint.

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/IEC 21122-1, *JPEG XS low-latency lightweight image coding system — Part 1: Core coding system*

3   Terms, definitions, symbols and abbreviated terms

3.1   Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 21122-1 and the following 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/

3.1.1
blanking codestream fragment
placeholder *codestream fragment* (3.1.8) representing blanking periods

3.1.2
horizontal blanking period
timespan expressed in units of the grid point sampling rate between the last *pixel* (3.1.22) of an image line — not being the last line of an image — and the first pixel of the next image line

3.1.3
vertical blanking period
timespan in units of the grid point sampling rate between the last line of an image [including the *horizontal blanking periods* (3.1.2)] and the first line of the next image

3.1.4
buffer model
combination of a *decoder model* (3.1.12) and a *channel model* (3.1.6) whose behaviour can be defined by a set of parameters

3.1.5
buffer model instance
specific configuration of a *buffer model* (3.1.4) specified by the assignment of well-defined values to the buffer model parameters

3.1.6
channel model
model describing the temporal behaviour of the *transmission channel* (3.1.29) connecting an encoder and a decoder

3.1.7
coded codestream fragment
continuous sequence of bits in the codestream containing exactly one packet body and a well-defined number of packet headers, markers and marker segments

3.1.8
codestream fragment
either *coded codestream fragment* (3.1.7), or *blanking codestream fragment* (3.1.1)

3.1.9
code group
group of quantization indices in sign-magnitude representation representing a quantized *coefficient group* (3.1.10)

3.1.10
coefficient group
number of horizontally adjacent wavelet coefficients from the same band and precinct

3.1.11
cycle
single clock period of an encoder or decoder clocked implementation

3.1.12
decoder model
combination of a *decoder unit* (3.1.14) and a *decoder smoothing buffer* (3.1.13)

3.1.13
decoder smoothing buffer
memory buffer that is used to level out changes in the number of bits read by a *decoder unit* (3.1.14) per time unit

3.1.14
decoder unit
module reading a variable number of bits per time unit to generate decoded output *pixels* (3.1.22) with a fixed rate

3.1.15
decomposition level
set of wavelet coefficients resulting from a particular *level* (3.1.21) of recursive application of a wavelet transform

3.1.16
encoder model
combination of an *encoder unit* (3.1.18) and an *encoder smoothing buffer* (3.1.17)

3.1.17
encoder smoothing buffer
memory buffer that is used to level out changes in the number of bits generated by an *encoder unit* (3.1.18) per time unit

3.1.18
encoder unit
module transforming a sequence of input *pixels* (3.1.22) with constant rate into a conforming codestream, producing a bit sequence with variable number of bits generated per time unit

3.1.19
fill level
number of bits stored in the encoder or *decoder smoothing buffer* (3.1.13)

3.1.20
nominal bits per pixel value
mean number of bits allocated per encoded *pixel* (3.1.22) which is used to derive the *sublevel* (3.1.28) constraints by assuming an image with well-defined dimensions and frame rate derived from the *level* (3.1.21)

3.1.21
level
defined set of constraints on the amount of decoded *sampling grid points* (3.1.25) to be processed by an encoder or decoder, both in the spatial and time dimensions

Note 1 to entry: The same set of levels is defined for all profiles. Individual implementations may, within the specified constraints, support a different level for each supported profile.

3.1.22
pixel
position in the *sample grid* (3.1.24) that is populated by a sample value of at least one component

3.1.23
profile
specified subset of the codestream syntax together with admissible parameter values

3.1.24
sample grid
abstract coordinate system on which image sample values are positioned

3.1.25
sampling grid point
position on the *sample grid* (3.1.24), specified by integer horizontal and vertical offset relative to the origin of the sample grid

3.1.26
smoothing buffer unit
level (3.1.21) and *sublevel* (3.1.28) dependent number of bits by which the smoothing buffer size of the *decoder model* (3.1.12) is specified

3.1.27
start of transmission
SoT
time at which the *transmission channel* (3.1.29) starts transmission relative to the start of encoding of the first *codestream fragment* (3.1.8) of a codestream

3.1.28
sublevel
defined set of constraints on the amount of codestream bits to be processed by an encoder or decoder, per unit of time, per column, and per image

Note 1 to entry: The same set of sublevels is defined for all profiles. Individual implementations may, within the specified constraints, support a different sublevel for each supported profile.

3.1.29
transmission channel
facility transferring bits from a source entity to a target entity

3.1.30
transmission channel capacity
maximum number of bits per time unit that a *transmission channel* (3.1.29) can transfer from a source entity to a target entityConventions

3.2   Conformance language

ISO/IEC's use of verbal forms is detailed at:

https://www.iso.org/foreword-supplementary-information.html

The keyword "reserved" indicates a provision that is not specified at this time, shall not be used, and may be specified in the future. The keyword "forbidden" indicates "reserved" and in addition indicates that the provision will never be specified in the future.

3.3   Operators

NOTE   Many of the operators used in document are similar to those used in the C programming language.

3.3.1   Arithmetic operators

+   addition

−   subtraction (as a binary operator) or negation (as a unary prefix operator)

×   multiplication

/   division without truncation or rounding

3.3.2   Logical operators

|| logical OR

&&   logical AND

!   logical NOT

3.3.3   Relational operators

>   greater than

≥   greater than or equal to

<   less than

≤   less than or equal to

==   equal to

!=   not equal to

3.3.4   Precedence order of operators

Operators are listed in descending order of precedence. If several operators appear in the same line, they have equal precedence. When several operators of equal precedence appear at the same level in an

expression, evaluation proceeds according to the associativity of the operator either from right to left or from left to right.

Operators	Type of operation	Associativity
()	expression	left to right
[]	indexing of arrays	left to right
–	unary negation	
×, /	multiplication, division	left to right
+, –	addition and subtraction	left to right
< , >, ≤, ≥	relational	left to right
&	bitwise AND	left to right
		bitwise OR	left to right

3.3.5 Mathematical functions

$\lceil x \rceil$ ceil of x: returns the smallest integer that is greater than or equal to x

$\lfloor x \rfloor$ floor of x: returns the largest integer that is less than or equal to x

$|x|$ absolute value of x, $|x|$ equals –x for x < 0, otherwise x

$\mathrm{sign}(x)$ sign of x, 0 if x is 0, +1 if x is positive, −1 if x is negative

$\xi(t)$ step function $\xi(t) = \begin{cases} 1 & t \geq 0 \\ 0 & \text{otherwise} \end{cases}$

$\max_i(x_i)$ maximum of a sequence of numbers $[x_i]$ enumerated by the index i

4 Specifications

4.1 Symbols

$A = [a_1, a_2, \ldots, a_n]$ sequence of elements $a_1 – a_n$

$A||B$ concatenation of two sequences A and B

$C(i)$ codestream *i*

D_{c2d} number of clock cycles between the first bit written into the decoding smoothing buffer and the decoding start of the first codestream fragment of a stream of codestream fragments

$F_{first}(C(i))$ first codestream fragment of codestream C(i)

$F_{last}(C(i))$ last codestream fragment of codestream C(i)

H_f height of the image in sampling grid points

H_{max}	maximum picture height in sampling grid points
L_{max}	maximum number of sampling grid points per image
$l_{enc}(t)$	fill level of the encoding smoothing buffer in bits at the end of cycle t
$l_{dec}(t)$	fill level of the decoding smoothing buffer in bits at the end of cycle t
$l_{enc,max}$	capacity in bits of the encoding smoothing buffer
$l_{dec,max}$	capacity in bits of the decoding smoothing buffer
$\tilde{l}_{dec}(t)$	number of bits that can be read from the decoding smoothing buffer in cycle t
$l_{sum}(t)$	sum of encoder and decoder smoothing buffer fill level in bits at cycle t
$\mathbb{N}$	all integer numbers being strictly larger than zero
$\mathbb{N}_0$	all integer numbers being greater than or equal to zero
$N_{b,x}$	size of the horizontal blanking line in sampling grid point clock periods
$N_{b,y}$	size of the vertical blanking period in sampling grid lines
N_{bpp}	nominal number of bits allocated per pixel for compression
N_c	number of components in an image
$N_{cg}(f)$	number of coefficient groups within codestream fragment f
$N_{cg,hz}$	number of coefficient groups associated to a codestream fragment representing a horizontal blanking period
$N_{cg,vt}$	number of coefficient groups associated to a codestream fragment representing a vertical blanking period
$N_f(i)$	number of codestream fragments within a codestream i
N_g	number of coefficients in a code group
$N_{L,y}$	number of vertical decomposition levels
$N_{p,x}$	number of precincts per sampling grid line
$N_{p,y}$	number of precincts per sampling grid column
N_{sbu}	number of decoder smoothing buffer units for a given profile
$\mathbb{Q}$	set of rational numbers
$r_{dec}(t)$	number of bits read and removed from the decoder smoothing buffer in clock cycle t
R_{trans}	transmission channel capacity, expressed in bits per cycle (having a duration of T)
$R_{t,max}(l_m, l_s)$	maximum admissible encoded throughput in bits per second for a given level
$R_{s,max}$	max grid point sample rate (in samples per second) at decoder output
$S_{bits}(f)$	number of bits forming the codestream fragment f

$S_{c,max}$	targeted maximum number of bytes of an encoded codestream
$S_{sbu}(l_m, l_s)$	size of the smoothing buffer unit in bytes for level l_m and sublevel l_s
$S_{sbo}(p)$	smoothing buffer increment in bits for a profile p
$S_{sl,max}(l_m, l_s)$	maximum size of an encoded codestream in bytes of level l_m and sublevel l_s
$s_x[i]$	subsampling factor of component i in horizontal direction
$s_y[i]$	subsampling factor of component i in vertical direction
T_{enc}	clock period defining the frequency by which code groups are processed by an encoder
T_{dec}	clock period defining the frequency by which code groups are processed by a decoder
$t_{enc,write}(f)$	timestamp in cycles at which the codestream fragment f is written to the encoder smoothing buffer
$t_{dec,start}(f)$	timestamp in cycles at which decoder starts decoding codestream fragment f
$t_{dec,read}(f)$	timestamp in cycles at which codestream fragment f is removed from the decoder smoothing buffer
Tbmd	buffer model type
$W_c[i]$	width of component i in samples
$W_{c,max}$	maximum column width in sampling grid points for a given profile
$w_{dec}(t)$	number of bits written into the decoder smoothing buffer in clock cycle t
W_f	width of the image in sampling grid points
W_{max}	maximum picture width in sampling grid points

4.2 Abbreviated terms

bpp	bits per pixel
DWT	discrete wavelet transform
IDWT	inverse discrete wavelet transform
RCT	reversible colour transform
IRCT	inverse reversible colour transform

4.3 General provisions

For a concrete application, only a subset of the codestream syntax specified in ISO/IEC 21122-1 is needed. Profiles as specified in 3.1.23 define corresponding interoperability points for those applications. In addition to profiles, levels (as specified in 3.1.21) and sublevels (specified in 3.1.28) limit the maximum throughput in the encoded (codestream) and decoded (pixel, spatial) domain. This allows creating cost-efficient implementations serving the needs of the desired applications.

Profiles, levels and sublevels shall be as specified in Annex A.

Keeping the end-to-end latency of an encoding-decoding chain under a given threshold is one of the main goals pursued by the methods defined in ISO/IEC 21122-1. To reach this goal, the definition of buffer models is necessary, consisting of a decoder model and a transmission channel model. The interaction

of a hypothetical reference decoder including its decoding smoothing buffer with a constant bitrate channel feeding this buffer shall be as specified in Annexes B and C. The size of this decoding smoothing buffer is specified in Annex A. Codestreams shall be formed such that this decoding smoothing buffer never overflows or underflows.

Buffer models are further discussed in Annex D. The buffer model provides encoders with the necessary information to generate codestreams that can be decoded by an arbitrary decoder implementation ensuring system interoperability.

In addition to the size of the decoder smoothing buffer, end-to-end latency also depends on the latency inherent to each processing step of the encoding-decoding chain whose methods are described in ISO/IEC 21122-1. To help implementers estimate the latency of their device, Annex E gives useful information on the minimum latency that can be achieved by the different methods described in ISO/IEC 21122-1.

5 Buffer model

5.1 General system block diagram

The JPEG XS coding system addresses applications where coded images are transferred from a source to a target, as shown in Figure 1. To this end, the encoder is compressing a continuous stream of input pixels into a sequence of bits. These bits are forwarded by means of a transmission channel to the decoder that decompresses the bits to produce a continuous stream of output pixels.
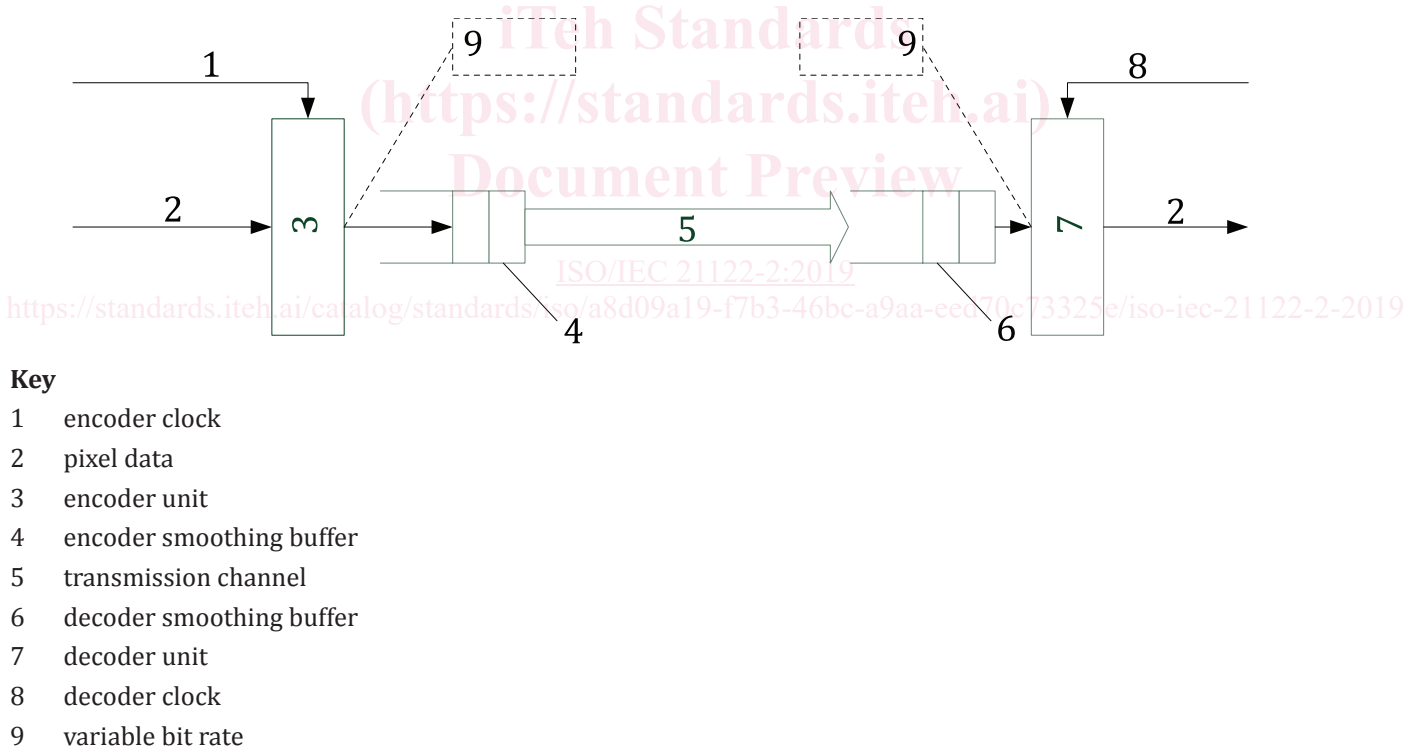
Key

1   encoder clock
2   pixel data
3   encoder unit
4   encoder smoothing buffer
5   transmission channel
6   decoder smoothing buffer
7   decoder unit
8   decoder clock
9   variable bit rate

Figure 1 — General system block diagram

The time instances at which the encoder has to process each pixel are determined by an encoding clock. Similarly, the time instances at which the decoder has to produce each output pixel are determined by a decoding clock. Both clocks are generated by the system.

NOTE   In implementations, these clocks can be the same or differ in both frequency and phase. The presented model is independent of whether clocks are synchronized or not.