INTERNATIONAL STANDARD

ISO/IEC 23001-11

First edition
2015-07-15

Information technology — MPEG systems technologies —

Part 11: Energy-efficient media consumption (green metadata)

Technologies de l'information — Technologies des systèmes MPEG — Partie 11: Consommation des supports éconergétiques (métadonnées vertes)

Reference number
ISO/IEC 23001-11:2015(E)

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

Contents

Page

Introduction

This part of ISO/IEC 23001 specifies the metadata (Green Metadata) that facilitates reduction of energy usage during media consumption as follows:

— the format of the metadata that enables reduced decoder power consumption;

— the format of the metadata that enables reduced display power consumption;

— the format of the metadata that enables media selection for joint decoder and display power reduction;

— the format of the metadata that enables quality recovery after low-power encoding.

This metadata facilitates reduced energy usage during media consumption without any degradation in the Quality of Experience (QoE). However, it is also possible to use this metadata to get larger energy savings, but at the expense of some QoE degradation.

Information technology — MPEG systems technologies —

Part 11:
Energy-efficient media consumption (green metadata)

1 Scope

This part of ISO/IEC 23001 specifies metadata for energy-efficient decoding, encoding, presentation, and selection of media.

The metadata for energy-efficient decoding specifies two sets of information: Complexity Metrics (CM) metadata and Decoding Operation Reduction Request (DOR-Req) metadata. A decoder uses CM metadata to vary operating frequency and thus reduce decoder power consumption. In a point-to-point video conferencing application, the remote encoder uses the DOR-Req metadata to modify the decoding complexity of the bitstream and thus reduce local decoder power consumption.

The metadata for energy-efficient encoding specifies a quality metric that is used by a decoder to reduce the quality loss from low-power encoding.

The metadata for energy-efficient presentation specifies RGB-component statistics and quality levels. A presentation subsystem uses this metadata to reduce power by adjusting display parameters, based on the statistics, to provide a desired quality level from those provided in the metadata.

The metadata for energy-efficient media selection specifies Decoder Operation Reduction Ratios (DOR-Ratios), RGB-component statistics and quality levels. The client in an adaptive streaming session uses this metadata to determine decoder and display power-saving characteristics of available video Representations and to select the Representation with the optimal quality for a given power-saving.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 13818-1:2013, *Information technology — Generic coding of moving pictures and associated audio information — Part 1: Systems*

ISO/IEC 14496-10, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 23001-10, *Information technology — MPEG systems technologies — Part 10: Carriage of Timed Metadata Metrics of Media in ISO Base Media File*

ISO/IEC 23009-1, *Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats*

ISO/IEC 23009-1:2015/Amd 2:2015, *Spatial relationship description, generalized URL parameters and other extensions*

ISO/IEC/TR 23009-3:2015, *Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 3: Implementation guidelines*

3 Terms, definitions, symbols, abbreviated terms and conventions

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

3.1 Terms and definitions

3.1.1
Adaptation Set
using the terms and definitions in ISO/IEC 23009-1, a set of interchangeable encoded versions of one or several media content components

3.1.2
Alpha-Point Deblocking Instance
APDI
using the notation, terms, and definitions in ISO/IEC 14496-10, a single filtering operation that produces either a single, filtered output p'_0 or a single, filtered output q'_0, where p'_0 and q'_0 are filtered samples across a 4x4 *block* edge

3.1.3
bitstream
using the terms and definitions in ISO/IEC 14496-10, a sequence of bits that forms the representation of coded pictures and associated data forming one or more coded video sequences

3.1.4
block
using the terms and definitions in ISO/IEC 14496-10, an MxN (M-column by N-row) array of samples or an MxN array of transform coefficients

3.1.5
byte
using the terms and definitions in ISO/IEC 14496-10, a sequence of 8 bits, written and read with the most significant bit on the left and the least significant bit on the right

3.1.6
chroma
using the terms and definitions in ISO/IEC 14496-10, an adjective specifying that a sample array or single sample is representing one of the two colour difference signals relating to the primary colours

3.1.7
chroma_format_idc
using the notation, terms and definitions in ISO/IEC 14496-10, specifies the chroma sampling relative to the luma sampling

3.1.8
decoded picture
using the terms and definitions in ISO/IEC 14496-10, a picture derived by decoding a coded picture

3.1.9
decoder
using the terms and definitions in ISO/IEC 14496-10, an embodiment of a decoding process

3.1.10
display process
using the terms and definitions in ISO/IEC 14496-10, a process that takes, as its input, the cropped decoded pictures that are the output of the decoding process

3.1.11
encoder
using the terms and definitions in ISO/IEC 14496-10, an embodiment of an encoding process

3.1.12
frame
using the terms and definitions in ISO/IEC 14496-10, an array of luma samples in monochrome format or an array of luma samples and two corresponding arrays of chroma samples in 4:2:0, 4:2:2 , and 4:4:4 colour format

3.1.13
informative
term used to refer to content provided in this Recommendation | International Standard that is not an integral part of this Recommendation | International Standard

3.1.14
intra coding
using the terms and definitions in ISO/IEC 14496-10, coding of a block, macroblock, slice or picture that uses intra prediction

3.1.15
luma
using the terms and definitions in ISO/IEC 14496-10, an adjective specifying that a sample array or single sample is representing the monochrome signal relating to the primary colours

3.1.16
macroblock
using the terms and definitions in ISO/IEC 14496-10, a 16x16 block of luma samples and two corresponding blocks of chroma samples of a picture that has three sample arrays, or a 16x16 block of samples of a monochrome picture or a picture that is coded using three separate colour planes

3.1.17
Media Presentation Description
MPD
using the terms and definitions in ISO/IEC 23009-1, a formalized description for a Media Presentation for the purpose of providing a streaming service

3.1.18
No-Quality-Loss Operating Point
NQLOP
metadata-enabled operating point associated with the largest display-power reduction that can be achieved without any quality loss (infinite PSNR)

3.1.19
non-zero macroblock
macroblock (3.1.16) containing at least one non-zero sample

3.1.20
note
term that is used to prefix *informative* (3.1.13) remarks (used exclusively in an informative context)

3.1.21
peak signal
maximum permissible *RGB component* (3.1.31) in a reconstructed frame

Note 1 to entry: For 8-bit video, peak signal is 255.

3.1.22
period
interval over which complexity-metrics metadata are applicable

3.1.23
Period
using the terms and definitions in ISO/IEC 23009-1, an interval of the Media Presentation, where a contiguous sequence of all Periods constitutes the Media Presentation

3.1.24
PicSizeInMbs
using the notation, terms and definitions in ISO/IEC 14496-10, a variable that is derived as the product of PicWidthInMbs and PicHeightInMbs

3.1.25
picture
using the terms and definitions in ISO/IEC 14496-10, a collective term for a field or a frame

3.1.26
pixel
smallest addressable element in an all-points addressable display device

3.1.27
prediction
using the terms and definitions in ISO/IEC 14496-10, an embodiment of the prediction process

3.1.28
reconstructed frames
frames obtained after applying RGB colour-space conversion and cropping to the specific *decoded picture* (3.1.8) or *pictures* (3.1.25) for which display power-reduction metadata are applicable

3.1.29
Representation
using the terms and definitions in ISO/IEC 23009-1, a collection and encapsulation of one or more media streams in a delivery format and associated with descriptive metadata

3.1.30
RGB colour space
colour space based on the red, green, and blue primaries

3.1.31
RGB component
single sample representing one of the three primary colours of the *RGB colour space* (3.1.30)

3.1.32
Segment
using the terms and definitions in ISO/IEC 23009-1, a unit of data associated with an HTTP-URL and optionally a byte range that are specified by an MPD

3.1.33
separate_colour_plane_flag
using the notation, terms, and definitions in ISO/IEC 14496-10, a flag that, when set, specifies that the three colour components of the 4:4:4 chroma format are coded separately

3.1.34
shall
term used to express mandatory requirements for conformance to this Recommendation | International Standard

3.1.35
should
term used to refer to behaviour of an implementation that is encouraged to be followed under anticipated ordinary circumstances, but is not a mandatory requirement for conformance to this Recommendation | International Standard

3.1.36
Six-Tap Filtering
STF
indicates a single application of the 6-tap filter, defined in ISO/IEC 14496-10, to generate a single filtered sample

3.1.37
source
using the terms and definitions in ISO/IEC 14496-10, a term used to describe some of the video material or some of its attributes before encoding

3.2 Symbols and abbreviated terms

For the purpose of this document, the symbols and abbreviated terms given in the following apply:

APDI	Alpha-Point Deblocking Instance
ASIC	Application Specific Integrated Circuit
AVC	Advanced Video Coding
CM	Complexity Metric
CMOS	Complementary Metal Oxide Semiconductor
CPU	Central processing Unit
DASH	Dynamic Adaptive Streaming over HTTP
DOR-Ratio	Decoding Operation Reduction Ratio
DOR-Req	Decoding Operation Reduction Request
DVFS	Dynamic Voltage Frequency Scaling
FS	Fresh Start
GP	Good Picture
MPD	Media Presentation Description
MSD	Mean Square Difference
MV	Motion Vector
NQLOP	No-Quality-Loss Operating Point
PSNR	Peak Signal to Noise Ratio
QoE	Quality of Experience
RBLL	Remaining Battery Life Level
RGB	Red, Green, Blue
SEI	Supplemental Enhancement Information
SP	Start Picture
STF	Six-Tap Filtering
XSD	Cross-Segment Decoding

3.3 Conventions

3.3.1 Arithmetic operators

+	Addition

-	Subtraction (as a two-argument operator) or negation (as a unary prefix operator)
*	Multiplication
x^y	Exponentiation
x/y	Division where no truncation or rounding is intended
$\frac{x}{y}$	Division where no truncation or rounding is intended
$\sum_{i=x}^{y} f(i)$	Summation of $f(i)$ with *i* taking all integer values from *x* up to and including *y*

3.3.2 Mathematical functions

Mathematical functions in this Technical Specification are defined as follows:

$$\mathrm{Abs}(x) = \begin{cases} -x, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{3-1}$$

$$\mathrm{Clip}(x) = \begin{cases} x, & x < 256 \\ 255, & \text{otherwise} \end{cases} \tag{3-2}$$

Floor(*x*) is the greatest integer less than or equal to *x* (3-3)

Log10(*x*) returns the base-10 logarithm of *x* (3-4)

Round(*x*) = Sign(*x*) * Floor(Abs(*x*) + 0.5) (3-5)

$$\mathrm{Sign}(x) = \begin{cases} -1, & x < 0 \\ 1, & x \leq 0 \end{cases} \tag{3-6}$$

4 Functional architecture (Informative)

This clause is informative and placed here to provide context.

4.1 Description of the functional architecture

Figure 1 shows the functional architecture utilizing Green Metadata in this Technical Specification. The media pre-processor is applied to analyse and to filter the content source and a video encoder is used to encode the content to a bitstream for delivery. The bitstream is delivered to the receiver and decoded by a video decoder with the output rendered on a presentation subsystem that implements a display process.

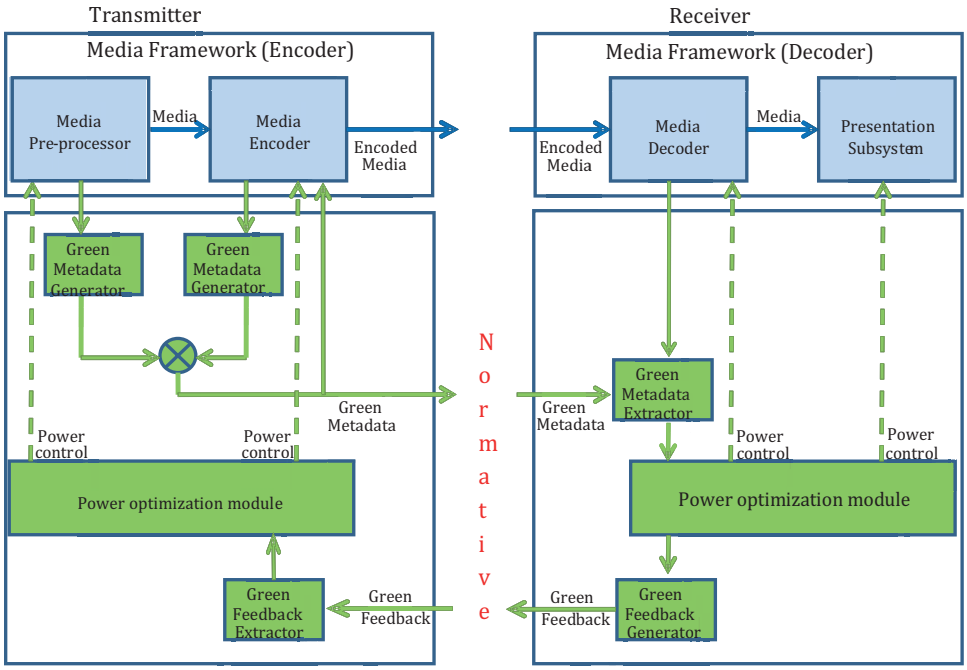

Figure 1 — Functional architecture

The Green Metadata is extracted from either the media encoder or the media pre-processor. In both cases, the Green Metadata is multiplexed or encapsulated in the conformant bitstream. Such Green Metadata is used at the receiver to reduce the power consumption for video decoding and presentation. The bitstream will be packetized and delivered to the receiver for decoding and presentation. At the receiver, the metadata extractor processes the packets and sends the Green Metadata to a power optimization module for efficient power control. For instance, the power optimization module interprets the Green Metadata and then applies appropriate operations to reduce the video decoder's power consumption when decoding the video and also to reduce the presentation subsystem's power consumption when rendering the video. In addition, the power-optimization module could collect receiver information, such as remaining battery capacity, and send it to the transmitter as green feedback to adapt the encoder operations for power-consumption reduction.

The normative aspect of this document is limited to the Green Metadata and Green Feedback in Figure 1.

4.2 Definition of components in the functional architecture

Green Metadata generator

— Generates metadata from either the video encoder or the content pre-processor.

Green Metadata extractor

— Interprets the bitstream syntax information and sends it to the power optimization module in the receiver.

Green feedback generator

— Generates feedback information for the transmitter.

— Communicates with the transmitter through a feedback channel, if available, for energy-efficient processing.

Green feedback extractor

— Receives the feedback from the receiver and sends it to the power optimization module in the transmitter.

Power optimization module in the transmitter

— Collects platform statistics such as the remaining battery capacity of the device in which the transmitter resides.

— Controls the operation of the Green Metadata generator, video encoder and content pre-processor.

— Processes green feedback.

Power optimization module in the receiver

— Processes the green-metadata information and applies appropriate operations for power-consumption control.

— Collects platform statistics such as remaining battery capacity of the device in which the receiver resides.

— Sends requests to Green feedback generator.

5 Decoder power reduction

5.1 General

Energy-efficient decoding is achieved with two types of metadata: Complexity Metrics (CMs) metadata and Decoding Operation Reduction Request (DOR-Req) metadata. A decoder may use CMs metadata to vary operating frequency and thus reduce decoder power consumption. In a point-to-point video conferencing application, the remote encoder may use the DOR-Req metadata to modify the decoding complexity of the bitstream and thus reduce local decoder power consumption.

5.2 Complexity metrics for decoder-power reduction

5.2.1 General

With respect to the functional architecture in Figure 1, the green-metadata generator provides CMs that indicate the picture-decoding complexity of an AVC bitstream to the decoder.

5.2.2 Syntax

The syntax for the CMs is as follows:

	Size (bits)	Descriptor
period_type	8	unsigned integer
if (period_type = = 2) {		
num_seconds	16	unsigned integer
}		
else if (period_type = = 3) {		
num_pictures	16	unsigned integer
}		
percent_non_zero_macroblocks	8	unsigned integer
percent_intra_coded_macroblocks	8	unsigned integer
percent_six_tap_filterings	8	unsigned integer
percent_alpha_point_deblocking_instances	8	unsigned integer

5.2.3 Signalling

SEI messages can be used to signal Green Metadata in an AVC stream. The Green Metadata SEI message payload type is specified in ISO/IEC 14496-10:2014/Amd. 2. The complete syntax of the Green Metadata SEI message payload is specified in Annex A.

The message containing the CMs is transmitted at the start of an upcoming period. The next message containing CMs will be transmitted at the start of the next upcoming period. Therefore, when the upcoming period is a picture or the interval up to the next I-slice, a message will be transmitted for each picture or interval, respectively. However, when the upcoming period is a specified time interval or a specified number of pictures, the associated message will be transmitted with the first picture in the time interval or with the first picture in the specified number of pictures.

5.2.4 Semantics

The semantics of various terms are defined below.

period_type – specifies the type of upcoming period over which the four complexity metrics are applicable and is defined in the following table.

Value	Description
0x00	complexity metrics are applicable to a single picture
0x01	complexity metrics are applicable to all pictures in decoding order, up to (but not including) the picture containing the next I slice
0x02	complexity metrics are applicable over a specified time interval in seconds
0x03	complexity metrics are applicable over a specified number of pictures counted in decoding order
0x04–0xFF	user-defined

num_seconds – when period_type is 2, num_seconds indicates the number of seconds over which the complexity metrics are applicable.

num_pictures – when period_type is 3, num_pictures specifies the number of pictures, counted in decoding order, over which the complexity metrics are applicable.