**ISO/IEC 23008-12:20xx/AMD 2:20xx(E)**

ISO/IEC JTC1/SC 29

Secretariat: JISC

**Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 12: Image File Format — Amendment 2: Support for tone map derived items and other improvements**

Preliminary WD stage

**Copyright notice**

This ISO document is a working draft or committee draft and is copyright-protected by ISO. While the reproduction of working drafts or committee drafts in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.

Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to ISO's member body in the country of the requester:

ISO copyright office

Case postale 56 • CH-1211 Geneva 20

Tel. + 41 22 749 01 11

Fax + 41 22 749 09 47

E-mail copyright@iso.org

Web www.iso.org

Reproduction for sales purposes may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

Contents

[Foreword iv](#_Toc149831930)

[1 Coding Constraints box related changes 1](#_Toc149831931)

[2 New tone-map derivation item 2](#_Toc149831932)

[6.6.2.4 Tone-map derivation 2](#_Toc149831933)

[6.6.2.4.1 Definition 2](#_Toc149831934)

[6.6.2.4.2 Syntax 3](#_Toc149831935)

[6.6.2.4.3 Semantics 4](#_Toc149831936)

[3 New constrained extents grid property 6](#_Toc149831937)

[6.5.37 Constrained Extents Grid Property 6](#_Toc149831938)

[4 New disparity adjustment information property 7](#_Toc149831939)

[6.5.37 Disparity adjustment information 7](#_Toc149831940)

[5 Unified identifier handling clarifications 7](#_Toc149831941)

[6 Overview images 8](#_Toc149831942)

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 documents 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](http://www.iso.org/directives) or [www.iec.ch/members\_experts/refdocs](http://www.iec.ch/members_experts/refdocs)).

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](http://www.iso.org/patents)) or the IEC list of patent declarations received (see [patents.iec.ch](https://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 the following URL:  [www.iso.org/iso/foreword.html](http://www.iso.org/iso/foreword.html). In the IEC, see [www.iec.ch/understanding-standards](https://www.iec.ch/understanding-standards)

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

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

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](https://www.iso.org/members.html) and [www.iec.ch/national-committees](http://www.iec.ch/national-committees).

Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 12: Image File Format — Amendment 2: Support for tone map derived items and other improvements

# Coding Constraints box related changes

In clause 7.2.3.4, renumber NOTE2 to NOTE3.

In clause 7.2.3.4, add the following text after NOTE1:

NOTE 2 When a track contains inter-predicted images and the value of all\_ref\_pics\_intra is equal to 0, it is possible for inter-predicted images to be derived from non-intra coded images. In such cases, derived specifications can suggest guidelines for the frequency of sync samples.

Add the following text as a new subclause after subclause 7.2.3.4:

7.2.3.5 Recommendations for CodingConstraintsBox

Encoding image sequences complying with the constraint that either all samples are sync samples or the all\_ref\_pics\_intra field in the CodingConstraintsBox can be set to one is suggested in the following cases:

1. For ensuring compatibility with players implementing codec brands in annexes B, E, L and M specified in an earlier edition of this document, which required all samples to be sync samples or to have the all\_ref\_pics\_intra field in the CodingConstraintsBox to be equal to 1.
2. In applications and usages where fast random access is essential, the constraint ensures random access to any image in the image sequence by decoding at most up to two images.
3. In image sequences that have an edit list causing backward playback, the constraint makes backward playback possible with an approach that each displayed image is obtained by decoding a compliant bitstream of up to two images.

In clause B.3.2, replace the following paragraph

For a track containing an HEVC image sequence, either all samples shall be sync samples or the all\_ref\_pics\_intra field in the CodingConstraintsBox specified in 7.2.3 shall be set to one.

with the following text

For a track containing an HEVC image sequence, either all samples should be sync samples or the all\_ref\_pics\_intra field in the CodingConstraintsBox specified in 7.2.3 should be set to one.

NOTE: Clause 7.2.3.5 contains recommendations that are important for backwards compatibility.

In clause E.3.2, replace the following paragraph

For a track containing an AVC image sequence, either all samples shall be sync samples or the all\_ref\_pics\_intra field in the CodingConstraintsBox specified in 7.2.3 shall be set to one.

with the following text

For a track containing an AVC image sequence, either all samples should be sync samples or the all\_ref\_pics\_intra field in the CodingConstraintsBox specified in 7.2.3 should be set to one.

NOTE: Clause 7.2.3.5 contains recommendations that are important for backwards compatibility.

In clause L.3.2, replace the following paragraph

For a track containing an VVC image sequence, either all samples shall be sync samples or the all\_ref\_pics\_intra field in the CodingConstraintsBox specified in ‎7.2.3 shall be set to one.

with the following text

For a track containing an VVC image sequence, either all samples should be sync samples or the all\_ref\_pics\_intra field in the CodingConstraintsBox specified in 7.2.3 should be set to one.

NOTE: Clause 7.2.3.5 contains recommendations that are important for backwards compatibility.

In clause M.3.2, replace the following paragraph

For a track containing an EVC image sequence, either all samples shall be sync samples or the all\_ref\_pics\_intra field in the CodingConstraintsBox specified in ‎7.2.3 shall be set to one.

with the following text

For a track containing an EVC image sequence, either all samples should be sync samples or the all\_ref\_pics\_intra field in the CodingConstraintsBox specified in 7.2.3 should be set to one.

NOTE: Clause 7.2.3.5 contains recommendations that are important for backwards compatibility.

# New tone-map derivation item

Add the following new subclause after subclause 6.6.2.3:

## Tone-map derivation

### 6.6.2.4.1 Definition

An item with an item\_type value of 'tmap' defines a derived image item whose reconstructed image is formed from one base input image and a secondary input image that will be referred to as a gain map input image.

The input images are given by the SingleItemTypeReferenceBox of type 'dimg' for this derived image item within the ItemReferenceBox. In the SingleItemTypeReferenceBox of type 'dimg', the value of from\_item\_ID identifies the derived image item of type 'tmap', the value of reference\_count shall be equal to 2, and the values of to\_item\_ID identify the input images, of which the first is considered as the base input image and the second as the gain map input image. The gain map input image may have different dimensions than the base as documented in ISO 21496-1 section 4.2.

Reconstruction is done by applying the gain map to the base image according to ISO 21496-1 section 6. As described in ISO 21496-1 section 6.5, the gain map may be scaled by a weight during application to adjust for local viewing conditions.

The base input image shall be associated with a 'colr' item property. This corresponds to the colorimetry metadata described in ISO 21496-1 section 5.3.1.

The gain map input image shall be associated with a 'colr' item property of type 'nclx' which indicates any transformations that the encoder has done to improve compression. In this item property, colour\_primaries and transfer\_characteristics shall be set to 2.

A 'tmap' derived item shall be associated with a 'colr' item property. This corresponds to the colorimetry metadata described in ISO 21496-1 section 5.3.2, which describes the colour properties of the reconstructed image if the gain map input item is fully applied according to ISO 21496-1 section 6.3.

When a 'tmap' derived item is the input to another derived item, that derived item shall treat the reconstructed image of the 'tmap' derived item as if the gain map has been fully applied and has the colour properties of the 'colr' item property associated with the 'tmap' derived item.

The base input image and the 'tmap' derived item should be associated with 'clli' item properties as appropriate to further document the optimal viewing conditions of each representation.

A 'tmap' derived item should be associated with a PixelInformationProperty item property. This property provides a hint to decoders on the approximate amount of colour resolution available after fully applying the gain map.

The number of channels in the gain map input item may be different than the number of channels in the channels field in the body of the 'tmap' derived item. If the gain map input item is single channel while the channels field is multi-channel, the gain map input item shall be treated as if it has three identical colour channels. If the gain map input item is multi-channel while the channels field is single channel, the channels field shall be treated as if it is multi-channel with identical values for all channels.

NOTE 1: The gain map input image should be marked as hidden by setting (flags & 1) equal to 0 in its infe entry.

NOTE 2: Backwards compatibility with parsers that do not support the tone-map derivation can be achieved by placing the base input image item and the 'tmap' derived item in an 'altr' entity group.

### Syntax

struct SignedRational {  
 signed int(32) numerator;  
 unsigned int(32) denominator;  
}

struct UnsignedRational {  
 unsigned int(32) numerator;  
 unsigned int(32) denominator;  
}

struct ToneMapChannelCommonDenominator {  
 int(32) gain\_map\_min\_numerator;  
 int(32) gain\_map\_max\_numerator;  
 unsigned int(32) gamma\_numerator;  
 int(32) base\_offset\_numerator;  
 int(32) alternate\_offset\_numerator;  
}

struct ToneMapChannel {  
 SignedRational gain\_map\_min;  
 SignedRational gain\_map\_max;  
 UnsignedRational gamma;  
 SignedRational base\_offset;  
 SignedRational alternate\_offset;  
}

aligned(8) class ToneMapImage {  
 unsigned int(8) version = 0;  
 if(version == 0) {  
 unsigned int(8) flags;  
 // temp/nonparsable variable  
 int channel\_count = (flags & 1)\*2 + 1;  
 Boolean use\_base\_colour\_space = (flags & 2) != 0;  
 Boolean backward\_direction = (flags & 4) != 0;  
 Boolean use\_common\_denominator = (flags & 8) != 0;  
 if (use\_common\_denominator) {  
 unsigned int(32) common\_denominator;  
 unsigned int(32) base\_hdr\_headroom\_numerator;  
 unsigned int(32) alternate\_hdr\_headroom\_numerator;  
 ToneMapChannelCommonDenominator channels[channel\_count];  
 }  
 else {  
 UnsignedRational base\_hdr\_headroom;  
 UnsignedRational alternate\_hdr\_headroom;  
 ToneMapChannel channels[channel\_count];  
 }  
 }  
}

### Semantics

version shall be equal to 0. Readers shall not process a ToneMapImage with an unrecognized version number.

(flags & 1) equal to 1 specifies that the channel count of the gain map per-channel metadata is 3, while a value of 0 specifies that the channel count is 1.

(flags & 2) equal to 2 specifies that the colour space of the base input shall be used as the gain map application space as described in ISO 21496-1 section 5.3.3. If not equal to 2, the colour space of the 'tmap' derived item is used as the gain map application space instead.

(flags & 4) equal to 0 specifies that the gain map is an inverted gain map and shall be applied with a weighting factor of -1 as expressed in ISO 21496-1 Annex B.

(flags & 8) equal to 8 specifies that all rational fields in the 'tmap' metadata share a common denominator.

common\_denominator expresses a common denominator shared by all metadata fields. common\_denominator shall not be 0.

channels specifies the per-channel metadata. If multi-channel, the order of the channels is R, G, B.

base\_hdr\_headroom\_numerator specifies the numerator of the Baseline HDR headroom (ISO 21496-1 section 5.2.9) when common\_denominator is used.

base\_hdr\_headroom specifies the numerator and denominator of the Baseline HDR headroom (ISO 21496-1 section 5.2.9) when common\_denominator is not used. base\_hdr\_headroom.denominator shall not be 0.

alternate\_hdr\_headroom\_numerator specifies the numerator of the Alternate HDR headroom (ISO 21496-1 section 5.2.10) when common\_denominator is used.

alternate\_hdr\_headroom specifies the numerator and denominator of the Alternate HDR headroom (ISO 21496-1 section 5.2.10) when common\_denominator is not used. alternate\_hdr\_headroom.denominator shall not be 0.

gain\_map\_min\_numerator specifies the numerator of the Per component gain map min value (ISO 21496-1 section 5.2.4) when common\_denominator is used.

gain\_map\_min specifies the numerator and denominator of the Per component gain map min value (ISO 21496-1 section 5.2.4) when common\_denominator is not used. gain\_map\_min.denominator shall not be 0.

gain\_map\_max\_numerator specifies the numerator of the Per component gain map max value (ISO 21496-1 section 5.2.5) when common\_denominator is used.

gain\_map\_max specifies the numerator and denominator of the Per component gain map max value (ISO 21496-1 section 5.2.5) when common\_denominator is not used. gain\_map\_max.denominator shall not be 0.

gamma\_numerator specifies the numerator of the Per component gamma value (ISO 21496-1 section 5.2.8) when common\_denominator is used.

gamma specifies the numerator and denominator of the Per component gamma value (ISO 21496-1 section 5.2.8) when common\_denominator is not used. gamma.denominator shall not be 0.

base\_offset\_numerator specifies the numerator of the Per component baseline offset (ISO 21496-1 section 5.2.6) when common\_denominator is used.

base\_offset specifies the numerator and denominator of the Per component baseline offset (ISO 21496-1 section 5.2.6) when common\_denominator is not used. base\_offset.denominator shall not be 0.

alternate\_offset\_numerator specifies the numerator of the Per component alternate offset (ISO 21496-1 section 5.2.7) when common\_denominator is used.

alternate\_offset specifies the numerator and denominator of the Per component alternate offset (ISO 21496-1 section 5.2.7) when common\_denominator is not used. alternate\_offset.denominator shall not be 0.

# New constrained extents grid property

Add the following new subclause after subclause 6.5.36:

### Constrained Extents Grid Property

#### Definition

Box type: 'cexg'

Property type: Descriptive item property

Container: ItemPropertyContainerBox

Mandatory (per item): No

Quantity (per item): At most one

The ConstrainedExtentsGridProperty descriptive item property indicates that each extent of the associated image item in the itemLocationBox is constrained to enclose data units of the item that are extractable as a contiguous byte range and are independently decodable and renderable as image tiles.

All data units or properties required to configure the decoder and decode an image tile shall be declared in the decoder configuration and initialization properties associated with the image item.

The reconstructed image of the associated image item is formed from one or more image tiles in a given grid order within a larger canvas.

The image tiles corresponding to the extents are inserted in row-major order, top-row first, left to right, in the order of the extents for the associated image item within the ItemLocationBox. The value of extent\_count within the ItemLocationBox shall be equal to (1+rows\_minus\_one)\*(1+columns\_minus\_one). All image tiles shall have exactly the same width and height, image\_tile\_width and image\_tile\_height. The reconstructed image is formed by tiling the image tiles into a grid with a column width equal to image\_tile\_width and a row height equal to image\_tile\_height, without gap or overlap. The grid of image tiles shall completely “cover” the reconstructed image of the associated image item, where image\_tile\_width\*columns is greater than or equal to image\_width and image\_tile\_height\*rows is greater than or equal to image\_height, where image\_width and image\_height are signalled in the ImageSpatialExtentsProperty associated with the image item.

#### Syntax

aligned(8) class ConstrainedExtentsGridProperty  
extends ItemFullProperty('cexg', version = 0, flags) {  
 // this is a temporary,non-parsable variable  
 unsigned int FieldLength = ((flags & 1) + 1) \* 16;  
 unsigned int(16) rows\_minus\_one;  
 unsigned int(16) columns\_minus\_one;  
 unsigned int(FieldLength ) image\_tile\_width;  
 unsigned int(FieldLength ) image\_tile\_height;  
}

#### Semantics

(flags & 1) equals to 0 specifies that the length of the fields image\_tile\_width and image\_tile\_height is 16 bits. (flags & 1) equals to 1 specifies that the length of the fields image\_tile\_width and image\_tile\_height is 32 bits. The values of flags greater than 1 are reserved.

image\_tile\_width, image\_tile\_height: specify respectively the width and height in pixels of the image tiles.

rows\_minus\_one, columns\_minus\_one: specify the number of rows of image tiles, and the number of image tiles per row. The value is one less than the number of rows or columns respectively. Image tiles enclosed in extents populate the top row first, followed by the second row and following rows, in the order of extents.

# New disparity adjustment information property

Add the following new subclause after subclause 6.5.37:

### Disparity adjustment information

#### Definition

|  |  |
| --- | --- |
| Box type: | 'dadj' |
| Property type: | Descriptive item property |
| Container: | ItemPropertyContainerBox |
| Mandatory (per associated identifier value): | No |
| Quantity (per associated identifier value): | At most one |
|  |  |

The disparity adjustment descriptive item property defines the suggested global disparity adjustment amount for a stereo pair.

This item property should only be associated with an item or entity group that describes a stereo pair. If associated with a stereo pair entity group, the FileTypeBox or the ExtendedTypeBox associated with the FileTypeBox shall indicate that the requirements of the 'unif' brand apply in the file.

EDITORS NOTE: See TuC on frame-packed stereo items

#### Syntax

aligned(8) class DisparityAdjustmentProperty  
extends ItemFullProperty('dadj', version = 0, flags = 0) {  
 signed int(32) disparity\_adjustment;  
}

#### Semantics

disparity\_adjustment in units of 1/10000 image widths. Positive values denote increased disparity with respect to parallel view direction.

# Unified identifier handling clarifications

Replace the word "item\_ID" in clauses 6.5.27.1, 6.5.18.1, 6.5.19.1, 6.5.35.1 with the phrase "identifier value"

Append the following text as a new paragraph in clauses 6.5.27.1, 6.5.18.1, 6.5.19.1, 6.5.35.1:

If this property is associated with an entity group, the FileTypeBox or the ExtendedTypeBox associated with the FileTypeBox shall indicate that the requirements of the 'unif' brand apply in the file.

# Overview images

Add the following new subclause after subclause 6.4.9:

**6.4.10 Overview images**

An overview image is a grid derived image item or a tiled pre-derived coded image item whose reconstructed image is formed from generating a lower resolution, ‘binned’ version of a base image item. The base image item is also a tiled image item. The tiling may be implemented using a feature of a specific codec, or by using a grid derived image item. When a grid derived image item is used, the input items to the grid define the tiles. Derived image items shall not be used as inputs to the image grid, due to the need for in place byte range accessing of content. Individual tiles shall be written contiguously in memory, thereby allowing access with a single read or write action.

A pre-defined coded image representing an overview image or an image item representing the base image that are tiled using a feature of a specific codec shall be stored in such a way that each extent identifies that data range corresponding to a tile, and shall be associated with a ConstrainedExtentsGridProperty indicating the constraint on the extents and describing the tiling grid.

An overview image shall be tiled using the same tiling scheme as the base image, i.e. if tiles in the base image are X by Y pixels, they are X by Y pixels in the overview image. In cases where the binned resolution results in a fractional, or incomplete tile at the end of a row (column), the last tile in a row (column) of tiles shall be padded with the value zero at the end of the row (column) to complete the last tile in the row (column). If necessary, the clean aperture transformative property ('clap') may be applied to crop padded rows and/or columns. The number of tiles in a row (column) of tiles is determined by dividing the width (height) of the overview image by the tile size in X (tile size in Y) and rounding up.

An image pyramid is generated by stacking a series of progressively binned overview images and creating an ImagePyramidEntityGroup. Each overview image is associated with the original full resolution base image, using a reference of type 'base'. The amount of binning of each overview image is indicated in the ImagePyramidEntityGroup. The image format of the overviews is the same as the base image item. i.e. number of bands, bit depth, color format, etc.

Region items associated with the base image may be replicated for individual overviews using an appropriate scaling associated with the level of binning for a particular overview and referenced to the specific overview.

NOTE 1: In this version of the document, the exact derivation process (approaches such as the sum, average, median, minimum, or maximum value of a binned region) used to produce an overview from the base image is left to the implementer.

NOTE 2: When removing (or modifying) an item that is marked as the base image of an overview image, the content of associated image overview items might need to be removed (or rewritten).

Add the following new subclause after subclause 6.8.9:

**6.8.10 Image Pyramid Entity Group**

**6.8.10.1 Definition**

Box Type: 'pymd'

Container: GroupListBox in a MetaBox at file level

Mandatory: No

Quantity: Zero or more

The ImagePyramidEntityGroup indicates a set of image items, formed as a base image item and a series of progressively binned overview image items, which together form an image pyramid.

Each overview image item has a reference to the original full resolution base image item, using a reference of type 'base'.

The ImagePyramidEntityGroup also provides overall information for the individual tiles inside the overview image items and base image item of the image pyramid.

The image format of the overview images shall be the same as the base image (i.e. number of bands, bit depth, color format, etc).

This entity group shall contain entity\_id values that point to a base image item and a set of overview image items and shall contain no entity\_id values that point to tracks. The entities shall be listed in the order of lowest resolution overview image item to the highest resolution overview image item, followed finally by the base image item of the image pyramid.

There may be multiple ImagePyramidEntityGroups in the same file with different group\_id values.

NOTE All the entities of a same ImagePyramidEntityGroup, or only some of them, can also be members of a same entity group of type 'prgr' if they are stored in the file for allowing a progressive refinement. They can also be members of a same entity group of type 'altr' if they are proposed by the content creator as alternatives to be displayed for players not supporting the ImagePyramidEntityGroup.

**6.8.10.2 Syntax**

aligned(8) class ImagePyramidEntityGroup  
extends EntityToGroupBox ('pymd', version = 0, flags = 0) {  
 unsigned int(16) tile\_size\_x;  
 unsigned int(16) tile\_size\_y;  
 for(i=0; i<num\_entities\_in\_group;i++) {  
 unsigned int(8) layer\_binning;  
 unsigned int(16) tiles\_in\_layer\_row\_minus1;  
 unsigned int(16) tiles\_in\_layer\_column\_minus1;  
 }  
}

**6.8.10.2 Semantics**

num\_entities\_in\_group: as defined for EntityToGroupBox. In addition, it also specifies the number of layers of the image pyramid.

tile\_size\_x, tile\_size\_y: indicate the size in pixels of a tile in the width and height dimension, respectively, for all layers of the image pyramid.

layer\_binning: Indicates for each layer of the pyramid the level of binning between the base image and the overview image. A 2x2 binning is defined to be a layer\_binning of 2, a 4x4 binning is defined to be 4, etc. The width and height for an overview image with layer\_binning of 2 is half the width and half the height of the base image, etc. A base image has a layer\_binning of 1.

tiles\_in\_layer\_row\_minus1, tiles\_in\_layer\_column\_minus1: Indicate the number of tiles minus one in a row and a column, respectively, of a specific layer. If the layer is represented by a grid derived image item, tiles\_in\_layer\_row\_minus1 is equal to rows\_minus\_one and tiles\_in\_layer\_column\_minus1 is equal to columns\_minus\_one. If the layer is represented by a tiled pre-derived coded image item with an ConstrainedExtentsGridProperty, then tiles\_in\_layer\_row\_minus1 is equal to rows\_minus\_one and tiles\_in\_layer\_column\_minus1 is equal to columns\_minus\_one.