ISO/IEC JTC 1/SC 29/WG 03 N0693

**ISO/IEC JTC 1/SC 29/WG 03  
MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title:** Technology under Consideration on ISO/IEC 23008-12

**Status:** Approved

**Date of document:** 2022-10-28

**Source:** ISO/IEC JTC 1/SC 29/WG 03

**No. of pages:** 8 (with cover page)

**Email of Convenor:** young.L @ samsung . com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 03 MPEG SYSTEMS**

**ISO/IEC JTC 1/SC 29/WG 03 N0693**

**October 2022, Mainz (DE)**

|  |  |
| --- | --- |
| **Title** | **Technology under Consideration on ISO/IEC 23008-12** |
| **Source** | **WG 03, MPEG Systems** |
| **Status** | **Approved** |
| **Serial Number** | **21972** |

# Abstract

This document collects following candidate technologies for the High Efficiency Image File Format (HEIF) (ISO/IEC 23008-12).

Table of Contents

[Abstract 1](#_Toc117776635)

[1. Region annotations for image sequence or video tracks 1](#_Toc117776780)

[**1** **Proposal (from m60304, MPEG#139, Issue#76)** 2](#_Toc117776781)

[**1.1** **Overview** 2](#_Toc117776782)

[**1.2** **Proposals** 2](#_Toc117776783)

[**2** **Proposal (m59508, MPEG#138)** 6](#_Toc117776784)

[**3** **Discussion** 9](#_Toc117776785)

[2. Matrix-based transformation for image items 9](#_Toc117776786)

[3. Signaling for pre-derived coded image items 9](#_Toc117776787)



# Region annotations for image sequence or video tracks

*[Ed. (FD). At MPEG#139, region extrapolation has been proposed as in section 1 below]*

*[Ed. (FD). It has been decided at MPEG#138 to move Region Annotations for image sequence and video tracks as proposed in m59508 to HEIF CDAM,* ***except the interpolate flag that may need further thoughts*** *(the yellow highlighted syntax and semantics (*[*comment on MPEG Gitlab*](http://mpegx.int-evry.fr/software/MPEG/Systems/FileFormat/HEIF/-/issues/69#note_60556)*).]*

1. **Proposal (from m60304, MPEG#139,** [**Issue#76**](http://mpegx.int-evry.fr/software/MPEG/Systems/FileFormat/HEIF/-/issues/76)**)**
   1. **Overview**

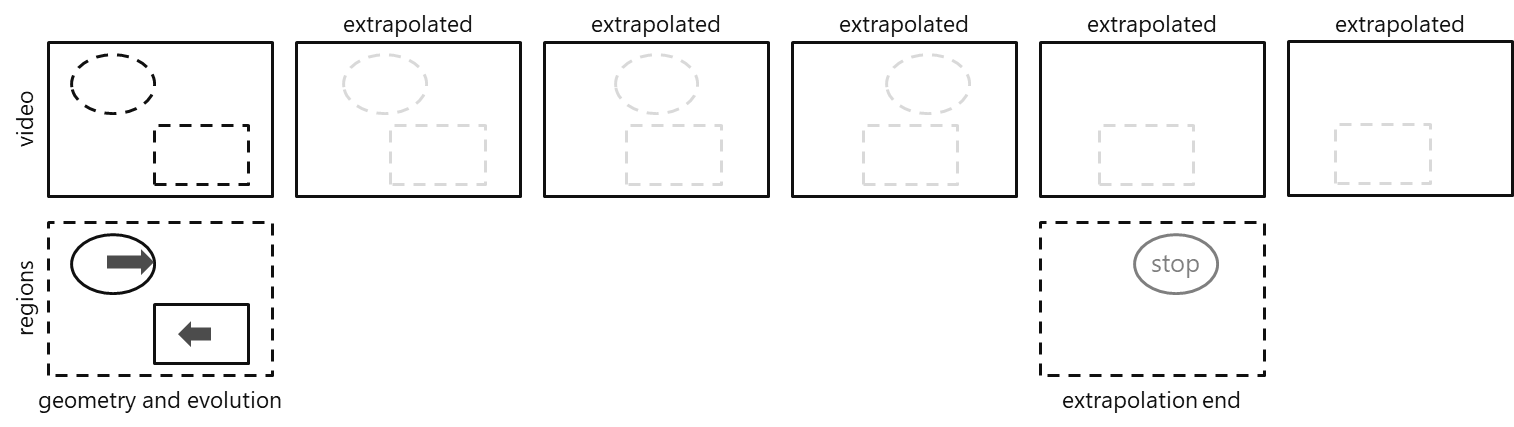


Figure : example region description for tracks using extrapolation

Figure 1 shows an example of describing several regions using extrapolation. The video track shown at the top contains two regions, an elliptic one and a rectangular one. The elliptic one is present in the four first samples of the video track and moves to the right of the image. The rectangular one is present in all the samples of the video track and moves to the left of the image.

The region track shown at the bottom describes these two regions. In a first sample, corresponding to the first sample of the video track, these two regions are described with their positions and sizes and the evolution of their respective positions and sizes. There are no region samples corresponding to the three following video samples. The region sample corresponding to the fifth video sample signals that the interpolation of the elliptic region ends.

* 1. **Proposals**

*Update the definition of a region track ( section 7.5.4.1) by adding the following paragraphs:*

The geometry of a region may be defined by specifying the shape, position and size of the region in a sample of the region track. The geometry of a region may also be defined as an initial geometry and its evolution over time by specifying the initial geometry of the region and its evolution in a sample of the region track.

The evolution of a region over time is optional. It can be represented by the evolution speed of some of its parameters inside the reference space. The evolution speed of the parameters is signaled using a scaling factor for increasing its precision. The parameters defining the evolution of a region depend on the geometry of the region as follows:

— When the geometry of a region is represented by a point, the evolution of the region is defined by the evolution of the position of this point.

— When the geometry of a region is represented by a rectangle or an ellipse, the evolution of the region is defined by the evolution of the position and the size of the rectangle or ellipse.

— When the geometry of a region is represented by a polygon or a polyline, the evolution of the region is defined by the evolution of the position of each point of the polygon or polyline. The number of points in the polygon or polyline doesn’t change.

— When the geometry of a region is represented by a mask, the evolution of the region is defined by the evolution of the position of the mask.

The evolution of a region stops when another sample contains a region with the same region identifier. The evolution of a region shall stop for each sync sample of the source track.

*Update the Sample format (section 7.5.4.2.1) with the following paragraph*

When the extrapolate flag is set to 1 for a region inside a sample of a region track, the region is an evolving region defined by an initial geometry and its evolution over time.

The value of each evolving parameter defining the geometry of the region at a given composition time *T* can be computed as follows:

where:

* *param0* is the initial value of the parameter as defined in the initial geometry of the region at time T0.
* *Δparam* is the evolution of the parameter as defined in the evolution of the region.
* *evolution\_scale* is a scaling factor for the evolution values equal to , where is the field\_size and is equal to ((RegionTrackConfigBox.field\_length\_size & 1) + 1) \* 16.
* *T0* is the composition time of the sample defining the evolving region.
* *ΔT* is the duration of the sample defining the evolving region.

*Update the syntax of Sample format (section 7.5.4.2.2) as follows*

aligned (8) class RegionSample {  
 unsigned int field\_size = ((RegionTrackConfigBox.field\_length\_size & 1) + 1) \* 16;  
// this is a temporary, non-parsable variable  
 unsigned int(32) region\_count;  
 for (r=0; r < region\_count; r++) {  
 unsigned int(32) region\_identifier;  
 unsigned int(8) geometry\_type;  
 unsigned int(1) extrapolate;  
 unsigned int(7) reserved;  
 if (geometry\_type == 0) {  
 // point  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 } }  
 else if (geometry\_type == 1) {  
 // rectangle  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) width;  
 unsigned int(field\_size) height;  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 signed int(field\_size) delta\_width;  
 signed int(field\_size) delta\_height;  
 }  
 }  
 else if (geometry\_type == 2) {  
 // ellipse  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) radius\_x;  
 unsigned int(field\_size) radius\_y;  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 signed int(field\_size) delta\_radius\_x;  
 signed int(field\_size) delta\_radius\_y;  
 }  
 }  
 else if (geometry\_type == 3 || geometry\_type == 6) {  
 // polygon or polyline  
 unsigned int(field size) point\_count;  
 for (i=0; i < point\_count; i++) {  
 signed int(field\_size) px;  
 signed int(field\_size) py;  
 }  
 if (extrapolate == 1) {  
 for (i=0; i < point\_count; i++) {  
 signed int(field\_size) delta\_px;  
 signed int(field\_size) delta\_py;  
 }  
 }  
 }  
 else if (geometry\_type == 4) {  
 // referenced mask  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) width;  
 unsigned int(field\_size) height;  
 unsigned int(field\_size) track\_mask\_idx;  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 }  
 }  
 else if (geometry\_type == 5) {  
 // inline mask  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) width;  
 unsigned int(field\_size) height;  
 unsigned int(8) mask\_coding\_method;  
 if (mask\_coding\_method != 0)  
 unsigned int(32) mask\_coding\_parameters;  
 bit(8) data[];  
 if (extrapolate == 1) {  
 signed int(field\_size) delta\_x;  
 signed int(field\_size) delta\_y;  
 }  
 }  
 else if (geometry\_type == 7) {  
 // empty region  
 }  
 }  
}

*Update the semantics of Sample format (section 7.5.4.2.3) with the following text:*

7: the region is an empty region used for signalling the end of the evolution of a previous region with the same region identifier.

Other values are reserved.

extrapolate is a flag indicating whether the geometry changes of the region are specified or not. When equal to 0, it indicates that no geometry changes are specified for the region. When equal to 1, it indicates that both the geometry and the geometry changes are specified for the region.

(…)

evolution\_scale is the scaling factor for the specification of the evolution values, equal to , where is the field\_size and is equal to ((RegionTrackConfigBox.field\_length\_size & 1) + 1) \* 16.

delta\_x, delta\_y specify, in 1/evolution\_scale units of the reference space, the evolution of the x and y fields for the region.

delta\_width, delta\_height specify, in 1/evolution\_scale units of the reference space the evolution of the width and height fields for the region.

delta\_radius\_x, delta\_radius\_y specify, in 1/evolution\_scale units of the reference space the evolution of the radius\_x and radius\_y fields for the region.

delta\_px, delta\_py specify, in 1/evolution\_scale units of the reference space the evolution of the px, py fields for a point of the region.

1. **Proposal (m59508, MPEG#138)**

*[Ed. (FD)This TuC only contains parts requiring further discussion (the interpolate flag in sample format for region tracks)]The HEIF CDAM defines the region track*

**X.X Region track and region annotations for an image sequence or video track**

**X.X.1 General**

A metadata track with a sample entry 'rgan' is a region track whose samples define one or more regions inside images carried in samples of an associated image sequence or video track (also denoted source track in the following).

*(… see HEIF CDAM for full text… )*

**X.X.3 Sample format**

**X.X.3.1 Definition**

This subclause defines the sample format for region track. A sample of a region track defines one or more regions.

**X.X.3.2 Syntax**

aligned (8) class RegionSample {  
 unsigned int field\_size = ((RegionTrackConfigBox.field\_length\_size & 1) + 1) \* 16;   
// this is a temporary, non-parsable variable  
 unsigned int(7)reserved;  
 unsigned int(1)interpolate;  
 unsigned int(16) region\_count;  
 for (r=0; r < region\_count; r++) {  
 unsigned int(32) region\_identifier;  
 unsigned int(8) geometry\_type;  
 if (geometry\_type == 0) {  
 // point  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 }  
 else if (geometry\_type == 1) {  
 // rectangle  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) width;  
 unsigned int(field\_size) height;  
 }  
 else if (geometry\_type == 2) {  
 // ellipse  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) radius\_x;  
 unsigned int(field\_size) radius\_y;  
 }  
 else if (geometry\_type == 3 || geometry\_type == 6) {  
 // polygon or polyline  
 unsigned int(field size) point\_count;  
 for (i=0; i < point\_count; i++) {  
 signed int(field\_size) px;  
 signed int(field\_size) py;  
 }  
 }  
 else if (geometry\_type == 4) {   
 // referenced mask  
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) width;  
 unsigned int(field\_size) height;  
 unsigned int(field\_size) track\_mask\_idx;  
 }  
 else if (geometry\_type == 5) {   
 // inline mask   
 signed int(field\_size) x;  
 signed int(field\_size) y;  
 unsigned int(field\_size) width;  
 unsigned int(field\_size) height;  
 unsigned int(8) mask\_coding\_method;   
 if (mask\_coding\_method != 0)   
 unsigned int(32) mask\_coding\_parameters;  
 bit(8) data[];  
 }  
 }  
}

**X.X.3.3 Semantics**

interpolate indicates the continuity in time of the successive samples. When true, the application may linearly interpolate values of the region geometries between the previous sample and the current sample. When false, there shall not be any interpolation of values between the previous and the current samples.

NOTE 1 When using interpolation, it is expected that the interpolated samples match the presentation time of the samples in the referenced source track. For instance, for each video sample of a video track, one interpolated region sample is calculated.

region\_count specifies the number of regions defined in the sample.

region\_identifier specifies the identifier of the region.

geometry\_type specifies the type of the geometry of a region. The following values for geometry\_type are defined:

0: the region is described as a point.

1: the region is described as a rectangle.

2: the region is described as an ellipse.

3: the region is described as a polygon.

4: the region is described as a mask defined in a referenced image item or in a sample of a referenced track.

5: the region is described as a mask defined inside the data of this sample.

6: the region is described as a polyline.

Other values are reserved.

x, y specify the coordinates of the point composing the region relatively to the reference space when its geometry is a point. x, y specify the top, left corner of the region relatively to the reference space when its geometry is a rectangle or a mask. x, y specify the centre of the region relatively to the reference space when its geometry is an ellipse. The value (x = 0, y = 0) represents the position of the top-left pixel in the reference space.

NOTE 2 Negative values for the x or y fields enable to specify points, top-left corners, and/or centres that are outside the image. This can be useful for updating region annotations during edition.

width, height specify, relatively to the reference space, the width and the height of the region when its geometry is a rectangle or a mask. When geometry\_type equals 4, the value 0 indicates that the corresponding width or height value is provided by the ImageSpatialExtentsProperty associated with the item containing the mask or the width or height of the track containing the mask. When geometry\_type does not equal 4, the value 0 is reserved.

radius\_x specifies, relatively to the reference space, the radius on the x-axis of the region when its geometry is an ellipse.

radius\_y specifies, relatively to the reference space, the radius on the y-axis of the region when its geometry is an ellipse.

point\_count is the number of points contained in a polygon or a polyline.

NOTE 3 A polygon specifying the geometry of a region is always closed and therefore there is no need to repeat the first point of the polygon as the ending point of the polygon.

px, py specify the coordinates of the points composing the polygon or the polyline relatively to the reference space. The value (px = 0, py = 0) represents the position of the top-left pixel in the reference space.

track\_mask\_idx specifies the index of the track reference of type 'mask' referring to the track from which to retrieve the mask to apply. The sample in that track from which mask data is retrieved is the one that is temporally aligned with the current sample in the source track or the nearest preceding one in the media presentation timeline. The first track reference has the index value 1; the value 0 is reserved.

mask\_coding\_method indicates the coding method applied on the mask contained in data. The following values are defined:

0: No mask encoding scheme is applied.

1: Mask is compressed with deflate() as defined in IETF RFC 1951.

Other values are reserved.

mask\_coding\_parameters indicates additional encoding parameters needed for successfully processing the coded mask data. When mask\_coding\_method is equal to 1, mask\_coding\_parameters indicates the number of bytes in the coded mask array data. The value of mask\_coding\_parameters is reserved when the value of mask\_coding\_method is greater than 1.

data contains the coded or uncompressed representation of a mask that contains the pixels for an inline mask in raster-scan order. Each pixel is represented using a single bit and 8 pixels are packed in one byte. Byte packing shall be in big-endian order. No padding shall be put at the end of each line if the width of the mask is not a multiple of 8 pixels. Only the last data byte shall be padded with bits set to 0.

1. **Discussion**

About the interpolate flag: The purpose is to avoid declaring a sample in the region track for each sample of the media track when regions are moving linearly between two positions. Imagine a sample A in the region track with a region at a starting position A and this region is moving linearly to the arrival position B nine samples later. Instead of declaring ten samples in the region track, you can only declare two samples, sample A with a duration corresponding to nine samples in the media track, followed by sample B providing the arrival position B. We should clarify that since the interpolate flag applies to all regions in the sample, the number of regions shall be the same in sample A and B.















# Matrix-based transformation for image items

*[[ Ed. (FD): MPEG#129: it was questioned:”* Should we also add ‘matrix’ as an image derivation in the HEIF? “. It was warned that “We would need to be clear about the meaning of outputs that don’t have horizontal and vertical sides; if that’s overlaid, the meaning is clear, but what if it’s supposed to be displayed?”*]]*

# Signaling for pre-derived coded image items

*Replace the clause 6.4.7 with the following text:*

**6.4.7** **Pre-derived coded images**

[Ed. (FD): In the following, differences with HEIF 2nd edition (w18310) are highlighted in blue]

If a coded image has been derived from others — for example, a composite HDR image derived from exposure-bracketed individual images, or a panorama derived from a set of images — then it shall be linked to those images by item references of type 'base'. Item references may be from the coded image to all images it derives from, or when unique IDs are used, from the coded image to all entity groups or images it derives from. When unique IDs are used, a to\_item\_ID value in the SingleItemTypeReferenceBox or SingleItemTypeReferenceBoxLarge is resolved to an item identifier whenever the embedding MetaBox contains an item with such identifier, and is resolved to an entity group identifier otherwise.

An image item including a 'base' item reference is referred to as a pre-derived coded image.

NOTE In this version of this document, the exact derivation process used to produce the image is not described.

[[Ed. (FD): At MPEG#129, it was commented that “The slight snag here is defining what it means when the entity group does NOT imply a single output (e.g. a slide show); what does pre-derivation mean? ]]

*Add the following clause as section 6.4.7.1:*

**6.4.7.1 Signaling of the derivation method for pre-derived coded image items**

A pre-derived coded image shall be linked to images it derives from by an item reference of type 'base' to the entity group containing all images the pre-derived coded images derives from. The grouping\_type of the EntityToGroupBox specifies the purpose of grouping and implicitly signals the type of the derivation operation which was applied to generate the pre-derived coded image.

[[Ed. (FM): At MPEG#126, it was commented that “we somehow need to indicate the derivation operation, rather than the nature of the input set”]]

[[Ed. (FD): At MPEG#129, it was commented that “We could allow a pre-derivation of the implied derivation of that entity group.”]]