**Agenda item:** 12.5

**Source:** Qualcomm Inc.

**Title: Carriage of HEIF Images in ITT4RT**

**Document for** Discussion andAgreement

# Introduction

Overlays are a key aspect of immersive conferencing as they allow super-imposing a wide range of content in the scene, such as slides, logos, and information screens. The overlay texture may be coming from video streams, images, or image sequences. The ITT4RT work item has a requirement to support images as sources for overlays.

In this contribution, we discuss the usage of the HEIF format as the format for image overlays.

# Background

## High Efficiency Image File (HEIF) Format

The HEIF format is defined in ISO/IEC 23008-12. It specifies the storage of images and image sequences and their metadata into ISOBMFF-based container files. HEIF defines a generic codec-agnostic structure for the storage of images, but also provides an instantiation based on H.265/HEVC codec.

HEIF follows the ISOBMFF design rules for the storage of media. As such, image sequences are stored in tracks and still images are stored as metadata items. In the former, the images are coded together, thus the coded images will have some decoding dependencies. In the latter, the images are coded independently from each other.

An ISOBMFF container file that stores images with HEVC encoding is identified by the “heic” or the ”hevc” brands correspondingly, which are found in the “ftyp” box at the beginning of the ISOBMFF box. The MIME types of these files is set to “image/heic” for still images and “image/heic-sequence” for image sequences.

The following figure shows a rough structure of the HEIF format that stores images and image sequences.



Example structure of HEIF container file

The images in a container may be assigned different roles. The following roles are identified:

* Cover image: the default image that is shown in the absence of other information on the display preference. Only one cover image may be defined in a HEIF
* Thumbnail image: a small resolution image corresponding to a master image
* Auxiliary image: an image that complements a master image, containing e.g. a depth map.
* Master image: a main image in the container file that is a full resolution.
* Hidden image: an image that is not intended to be displayed.
* Pre-computed derived image: a coded image that has been derived from other images
* Coded image: a coded representation of an image
* Derived image: is an image that has to be derived using some operations and a reference to other images in the container file. The defined operations include: cropping, rotation, mirroring, and composition.

Image sequences may only take master, auxiliary, or thumbnail image sequence roles.

The images are stored with their own metadata in the container file. The initialization information is stored in the sample entry as part the sample description box for image sequences. For still images, it is stored as part of the item property container box as item properties. This metadata contains information like dimensions of the image, codec initialization information, etc.

Additional metadata such as pixel aspect ratio, color information, and bit depth can be present in item property container boxes and reference the actual images using the “cdsc” reference type in the item reference box.

For image sequences, the specification defines playback control features to describe:

* Images that are not to be displayed
* Slideshow vs image collection
* Playback timing for slideshow
* Looping of the slideshow
* Transformations on the images

To facilitate access to a specific image in an image sequence, sample grouping is used to indicate the decoding dependencies of that image. The decoder doesn’t need to decode all images of an image sequence.

## RTP Payload Format for HEVC

The RTP payload format for HEVC is defined in [2]. It is currently supported by MTSI for the transport of video streams.

The payload format defines the usage of the RTP header when carrying HEVC payload, the packetization of the HEVC coded NAL units into RTP packets, the SDP description, and payload header extension mechanisms.

# Carriage of Images and Image Sequences

Both images and image sequences may be used as overlays or 360 backgrounds in the context of ITT4RT. The HEIF format is used as the source for the overlay. The transmission shall be HEVC compliant, i.e., the image items and image sequences shall be extracted from the HEIF source and transmitted using the HEVC payload format.

The ITT4RT-tx client shall indicate the following information to the receiver:

* That the stream is an image stream
* Whether the stream carries image sequences or still images or both.
* The number of still image items
* The number of images in an image sequence
* The transmission mode for the still images or image sequences. The transmission mode may be set to “coupled” to indicate the coupling of transmission and display, in which case the display time is always determined by the RTP timestamp. Alternatively, it may be “decoupled” to indicate that the transmission is independent from presentation, so that the presentation timing is provided separately to support use cases such as storage and looping.
* The display order of the images in a still image collection or image sequences, in case the transmission mode is set to “decoupled”.
* Image metadata for each of the image items or samples, including image dimensions, image role, etc.

To support this signaling, a new “image” SDP attribute is added to identify that the stream carries still images or an image sequence. The “image” attribute has the following ABNF syntax:

image\_attribute=”a=image:” pt SP transmission\_mode SP item\_count

transmission\_mode=”tmode=” (”coupled” / “decoupled” [“;” store] [“;” loop])

item\_count=”count=” 1\*DIGIT

metadata\_index=”meta=” 1\*DIGIT

store=”store”

loop=”loop”

The fields have the following semantics:

**pt**: the payload type used for the carriage of this image collection or image sequence. In the presence of multiple image sequences or a mix of image collections and image sequences, different payload types shall be used for the different sequences or image collection.

**transmission\_mode**: the transmission mode coupled, indicates that the overlay images will be continuously streamed (retransmitted) for as long the overlay is to be rendered. The RTP timestamp shall be used to determine the presentation time of the overlay image. The images are not required to be stored and no looping is performed. All images are expected shall be master images in this case. In case of “decoupled” transmission mode, the presentation of the images is overwritten by the image metadata. The RTP timestamps shall be used to determine the index of the images.

**item\_count**: provides the number of images in the corresponding image collection or image sequence.

**store**: for the decoupled transmission mode, the store flag tells the receiver whether to store the images for continuous presentation or not. This allows the transmission session to be much shorter than the actual presentation. In particular, a still image overlay is typically stored for presentation throughout the lifetime of the overlay.

**loop**: for the decoupled transmission mode, the loop flag indicates if a the image collection or image sequence shall be looped or whether the last image item in the collection or image sequence is to remain in display.

The carriage of the metadata for the image items in an image collection or image sequence is performed using the “image-metadata” attribute, which is defined as follows:

image-metadata=”a=image-metadata:” **pt** SP coded-metadata

Where coded-metadata is base64 coded image metadata for the corresponding image collection or image sequence.

The image metadata shall have the following format:

|  |  |
| --- | --- |
| image\_metadata(pt,image\_count) { |  |
| for(i=0;i<image\_count;i++) { |  |
| image\_properties() |  |
| display\_info() |  |
| extension\_info() |  |
| } |  |
| } |  |
| image\_properties() { |  |
| clap\_present\_flag | 1 |
| irot\_present\_flag | 1 |
| imir\_present\_flag | 1 |
| other\_properties() | 1 |
| reserved | 4 |
| if (clap) CleanApertureBox() |  |
| if(irot) ImageRotation() |  |
| if(imir) ImageMirror() |  |
| other\_properties() |  |
| } |  |
| display\_info() { |  |
| Order | u(8) |
| Duration | u(16) |
| } |  |
| other\_properties() { |  |
| Property\_count |  |
| for (i=1;i<=property\_count;i++) { |  |
| 4cc\_code |  |
| ItemProperty() |  |
| } |  |

All image properties shall be formatted according to their definition in [1].

The display\_info provides information on the display order of each image in the image collection or sequence. An order of 0 means the image is not to be rendered. This might be the case for auxiliary or hidden images. The duration indicates for how long the image is to be rendered in units of a 90kHz clock.

The metadata information shall be base64 encoded and provided as part of the image-metadata attribute for each image collection or image sequence independently.

Note that coupled transmission mode may be suitable for a live overlay stream, e.g. a live slide presentation. The decoupled transmission mode is suitable for preset content, such as a timed slideshow.

An ITT4RT-Rx client that doesn’t support static images will remove the image attribute. In such case, the ITT4RT-Tx client should transmit the static image as a regular continuous video stream that conforms to the MTSI requirements for an HEVC compressed video stream. If that is not possible, the ITT4RT-Tx client shall revise its offer to remove that media line.

An ITT4RT-Tx client that supports the image attribute but does not support the decoupled mode or cannot provide the requested storage functionality shall reply with the mode set to “coupled” in the answer. The ITT4RT-Tx client shall support falling back to the coupled mode.

An ITT4RT-Rx client that supports static images shall support the Main profile and the Main Still Picture profile of HEVC.

# Example SDP

The following table shows an example SDP for the carriage of image collections and image sequences:

|  |
| --- |
| .  .  .  m=video 49170 RTP/AVP 98,99  a=rtpmap:98 H265/90000  a=fmtp:98 profile-id=1;sprop-vps=<video parameter sets data>  a=**image**:98 tmode=coupled count=14  a=**image-metadata**:98 coded-metadata=<base64 coded metadata> a=rtpmap:99 H265/90000  a=fmtp:99 profile-id=1;sprop-vps=<video parameter sets data>  a=**image**:99 tmode=decoupled;store=1;loop=1 count=6 a=**image-metadata**:99 coded-metadata=<base64 coded metadata>  .  .  . |

# Proposal

We propose to agree the content of this document into the PD and to start a CR with the content of section 3 to complete the specification of the image overlays.

# References

[1] ISO/IEC 23008-12, MPEG systems Technologies – Part 12: Image File Format

[2] IETF RFC 7798, RTP Payload Format for High Efficiency Video Coding (HEVC)