

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2013/W13961
November 2013, Geneva, Switzerland

Title WD of carriage of still image and image sequences
Status Working draft
Authors Vinod Kumar Malamal Vadakital, David Singer
Group Systems/File format

1. Abstract

This working draft specifies file format support for the storage and presentation of HEVC coded still images and image sequences.

2. Introduction

HEVC still images, from a file format's perspective, are pictures that use the HEVC intra coding mode for encoding. They are typically coded at very high quality.

Image sequences can be efficiently coded using an HEVC video coder - either in all-intra mode when minimum decoder delay is required, or in inter predicted mode when high coding efficiency is preferred. They differ from a video sequence in that there need not be a presentation time associated with the frames of a coded image sequences. Examples of typical image sequences include:

1. Photo bursts
2. Focal stacks
3. Exposure stacks
4. Multi-spectral image sequences
5. Short animations similar to animated GIFs

Considering the requirements for image sequences coded using HEVC, any image of the coded image sequence should be accessible for manipulation with a low decoding delay. Examples of prediction structures to enable such a constraint are shown in Figures 1 to 4. In these figures, R indicates a picture that may be used as inter prediction reference for other pictures, and O indicates an output picture (with `pic_output_flag` equal to 1 in the HEVC coding specification context).



Figure 1: Using all intra coding.

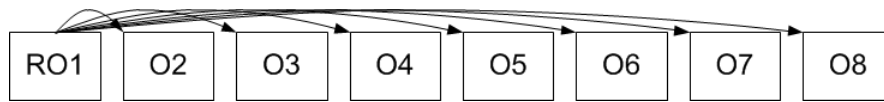


Figure 2: Using predictive coding with a single reference picture (the first picture in the sequence).

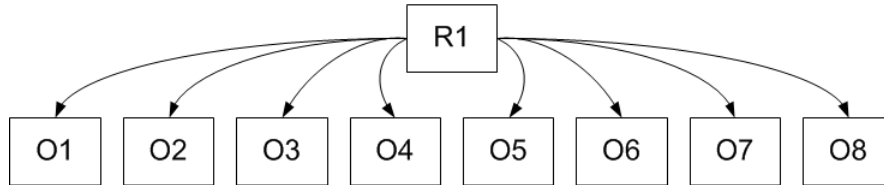


Figure 3: Using a common non-output reference picture.

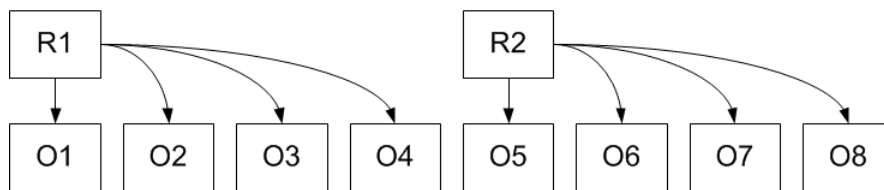


Figure 4: Using multiple non-output reference pictures.

Image sequences can be used either for simple storage purposes, or used as a part of a composite presentation of ISOBMFF compatible media.

3. Storage of a single still image into a file

TBD.

4. The still image/image sequence track

In [1], a track is defined to be a sequence of samples. Every sample in a track possess their own unique decode and presentation times. Storage of HEVC coded samples use sample entries of type 'hvc1' or 'hev1' for any HEVC specific decoder initializations. Since HEVC coded still images and image sequences follow the same encoder/decoder specifications, it is desirable to harmonize the file format structures as much as possible. Hence, the track structure is also used to encapsulate HEVC coded still images and image sequences. The choice of such a design allows using HEVC intra coded pictures as part of both a still picture track as well as a video track. The design also allows easy conversion between video and still image tracks, as well as untimed image sequences and video. Figure 5 illustrate the use of tracks to carry still images and image sequences.

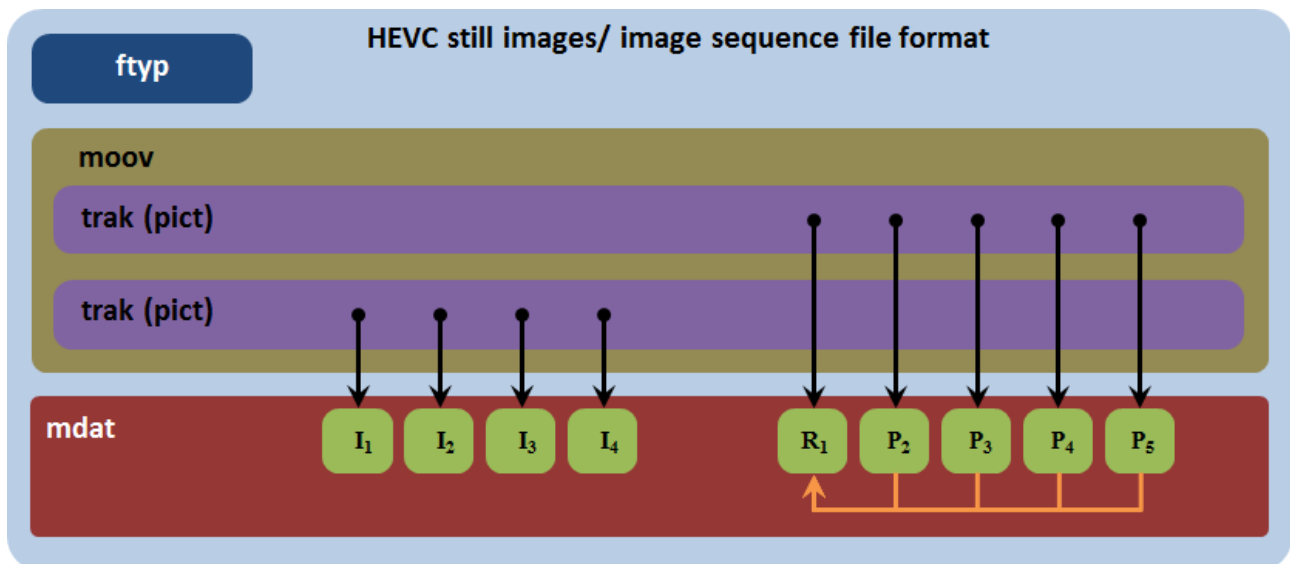


Figure 5: The HEVC still images/image sequence file format.

Handler type 'pict' is used to distinguish still images from other type of media. A handler type different from timed media makes it easy (e.g. for a browser) to conclude which entity should play a file. For example, if the file contains timed media ('soun', 'vide'), a media player may be launched, whereas if the file contains only a still picture, the browser may itself parse the file and display the image as part of the containing web page. The definition, syntax, and semantics of the SampleDescriptionBox is provided below.

Definition:

Box Types: 'stsd'
 Container: Sample Table Box ('stbl')
 Mandatory: Yes
 Quantity: Exactly one

Syntax:

```
aligned(8) abstract class SampleEntry (unsigned int(32) format)
extends Box(format){
    const unsigned int(8)[6] reserved = 0;
    unsigned int(16) data_reference_index;
}

class VisualSampleEntry(codingname) extends SampleEntry (codingname){
    unsigned int(16) pre_defined = 0;
    const unsigned int(16) reserved = 0;
    unsigned int(32)[3] pre_defined = 0;
    unsigned int(16) width;
    unsigned int(16) height;
    template unsigned int(32) horizresolution = 0x00480000; // 72 dpi
    template unsigned int(32) vertresolution = 0x00480000; // 72 dpi
    const unsigned int(32) reserved = 0;
    template unsigned int(16) frame_count = 1;
    string[32] compressorname;
    template unsigned int(16) depth = 0x0018;
    int(16) pre_defined = -1;
    // other boxes from derived specifications
    CleanApertureBox clap; // optional
    PixelAspectRatioBox pasp; // optional
}
```

```

aligned(8) class SampleDescriptionBox (unsigned int(32) handler_type)
extends FullBox('std', 0, 0){
    int i ;
    unsigned int(32) entry_count;
    for (i = 1 ; i <= entry_count ; i++){
        switch (handler_type){
            case 'soun': // for audio tracks
                AudioSampleEntry();
                break;
            case 'vide': // for video tracks
                VisualSampleEntry();
                break;
            case 'pict': // for Image tracks
                VisualSampleEntry();
                break;
            case 'hint': // Hint track
                HintSampleEntry();
                break;
            case 'meta': // Metadata track
                MetadataSampleEntry();
                break;
        }
    }
}

```

Semantics:

The semantics of all fields in the boxes remain unchanged from their previous definitions in [1].

When a track is used for HEVC still pictures using the 'pict' handler type, the `VisualSampleEntry` class is extended by `HEVCSampleEntry()` in exactly the same way as is done for a HEVC coded media using a 'vide' handler.

5. The TimeToSampleBox

The `TimeToSampleBox` is a mandatory box contained within the `TrackBox`. This box provides the decode time of samples in a compact manner. Entries in this table provide information about the number of consecutive samples having the same delta time, where the delta time $d(.)$ for a sample n is computed as $d(n) = dt(n) - dt(n-1)$; $dt(n)$ is the decode time of the n^{th} sample.

A still image/image sequence file composer may choose to either set the delta time to some constant value, or not code any delta value at all by setting the `entry_count` field in the `TimeToSampleBox` to zero. A still image/image sequence file reader, upon encountering a track using a handler of type 'pict', coded with a constant value delta time in its `TimeToSampleBox`, may either choose to ignore the decode timings, or use the decode timings to present the still images/image sequence as a slide show (depending on the use case).

6. Signalling of constraints

Image sequence requires additional signalling for constraints such as prediction structures. To enable such signalling, `HEVCSampleEntry` (an instantiation of the `VisualSampleEntry`), is extended by inserting an additional box that signals any such constraints. The new box is called the `CodingConstraintBox` with the fourCC 'ccst'. The `CodingConstraintBox` box carries information helpful for a file reader to adjust its operation when decoding HEVC image sequences. It includes indications such as whether all pictures or reference pictures within the track are intra coded or not. The presence of this box is mandated for tracks containing HEVC image sequences, but it can also be optionally inserted into tracks containing video sequences.

Definition:

Box Types:	'hvc1', 'hev1'
Container:	Sample description box ('std')
Mandatory:	An 'hvc1' or 'hev1' sample entry is mandatory
Quantity:	One or more sample entries may be present

Syntax:

```

class HEVCSampleEntry() extends VisualSampleEntry ('hvc1' or hev1'){
    HEVCConfigurationBox config;
    MPEG4BitRateBox ();           // optional
    MPEG4ExtensionDescriptorsBox (); // optional
    CodingConstraintsBox ();       // Mandatory for image sequences
    extra_boxes boxes;           ..// optional
}

class CodingConstraintsBox extends FullBox('ccst', version = 0, 0){
    unsigned int (1) AllReferencePicturesIntra;
    unsigned int(31) ReservedFlags;
    Box(); // zero or more reserved boxes
}

```

Semantics:

AllReferencePicturesIntra: This flag when set to one indicates the restriction that samples that are not sync samples, if any, are predicted only from sync samples. In other words, if there are inter predicted pictures in the track, then these pictures are all predicted from intra coded pictures. For a track containing an HEVC image sequence, either all samples shall be sync samples or the AllReferencePictureIntra shall be set to one.

7. Metadata for still images and image sequences

TBD

8. References

- [1] ISO/IEC 14496-12, Information technology — Coding of audio-visual objects — Part 12 -ISO base media file format.