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

ISO/IEC JTC1/SC29/WG11 MPEG2015/M37253
October 2015, Geneva, Switzerland

Source **Telecom ParisTech, Canon Research Centre France**
Status **For consideration at the 113th MPEG Meeting**
Title **Comments on LHEVC tracks layering**
Author Jean Le Feuvre, Cyril Concolato, Franck Denoual, Frédéric Mazé, Eric Nassor

1 Introduction

During MPEG 112, concerns were raised that a single layer per track design would complicate the job of the file writer and may impact the file size by duplicating the different tables. The minutes of the File Format Break-out Group also indicate that "*It seems that not using extractors may lead to a slight problem – in temporal scalability one would naturally need a sample in the higher layer.*"
In this contribution we review the implicit reconstruction process as documented in the DIS text.

2 Implicit reconstruction process of access units

The implicit AU reconstruction process allows reconstructing access units without extractors.
In such case, the file reader is in charge of:

- locating the proper layers required for the target operation point ('oinf' sample group)
- checking whether the track contains these layers for a given sample decode time ('linf' sample group)
- reconstructing the AU of a given decode time DTS, gathering all samples in identified tracks with the same DTS, if any
- reorganizing (mostly re-ordering) the NALUs per increasing nuh_layer_id and then temporal sublayer IDs

This reconstruction method works for all cases, including:

- all layers in a single track
- all layers in a dedicated track
- mix of layers in the same tracks, other layers dedicated tracks
- splitting of temporal layers in different tracks (the 'stts' table will obviously need adjustment).

Filtering of layers is quite simple:

- since all tracks containing layers of the same base shall use the same timescale, matching the samples across tracks is a simple matter of checking 'stts' entries (note that this process is also required for extractors resolution).
- Walking through NALUs and selecting them based on their layer ID (note that this process is also required for extractors resolution).

As shown in m36560 extractors can be quite costly, whether for tiles or for L-HEVC. Given this, we wonder why extractors are actually needed (remember MPEG "*one tool one functionality*"). The specification should either provide strong use cases for extractors that cannot be fulfilled with implicit AU reconstruction, and restrict the use of extractors to these use cases, or maybe extractors for HEVC should simply be dropped since extractor cookbooks are as computationally costly as implicit reconstruction.

3 Indicating track hierarchy with implicit reconstruction process

There are cases where identifying hierarchical dependencies between tracks containing a layered coded media is useful; especially in DASH delivery, the mapping of track references to Representation's dependencyId. The drawback with implicit AU reconstruction is that no track dependencies are indicated, since 'scal' dependencies are used for extractors reconstruction only. Using extractors only for indicating such dependencies is quite costly (12+ bytes per extractor).

When hierarchy is well defined among LHEVC tracks, we suggest allowing describing them through a 'dpnd' track reference to indicate track containing the lower layers for a given LHEVC track, in their order of dependencies. If extractors were to be removed for LHEVC, obviously the 'scal' track reference should be reused for this purpose.

4 Further simplifications on layering

Given the high number of possible combinations for storing LHEVC in ISOBMFF and previous discussions at MPEG 110, 111 and 112, we suggest restricting the storage of HEVC layers by allowing only the following combinations:

- all layers including the base HEVC in a single track
- each layer in a dedicated track, or multiple tracks when tile tracks are used, including its temporal sublayers.
- each layer in a dedicated track, or multiple tracks when tile tracks are used, and each temporal sublayer in a dedicated track.

The advantage of such limitations is that it keeps track dependencies simple.

These rules could either be in the core specification, or defined by a new brand "lhev".

5 Conclusion

We suggest MPEG to issue a study text for ISO/IEC 14496-15 taking the suggested modifications into account.