|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  37th Meeting: Geneva, CH, 4–10 October 2019 | Document: JCTVC-AK0028-v1 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Sub-layer picture rates SEI message for HEVC / AVC** | | |
| *Status:* | Input document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Sychev Maxim | Email: | Sychev.Maxim@huawei.com |
| *Source:* | Huawei Technologies Co., LTD | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This contribution presents signalling of mapping sub-layers to picture rates in video coding. For signaling of these parameters the Supplemental Enhancement Information (SEI) design is proposed to use. More specifically, the signaling of the table is proposed that allows mapping of Temporal IDs (TID) of sub-layers with corresponded picture rates.

# Introduction

A video of a given number of pictures may be partitioned into one or multiple coded video sequences (CVS). Each CVS can be decoded independently from other coded video sequences that may be contained in the same bitstream. The output order of the pictures is the order in which the pictures have been generated (e.g. recorded or rendered) and should be output for display in order of the original input sequence for the intended visual impression. In the specification, the output order is represented by the picture order count (POC), which uniquely identifies a picture within the CVS. POC distance between two successive pictures does not need to be constant throughout the CVS. It is only required to be strictly increasing along the output order.

A temporal layer concept in VVC is inherited from temporal scalability in H.265/HEVC and in H.264/AVC Annex G and available in the base specification for adaptation to variable channel bandwidth and for error resilience needs. Each picture has an associated temporal level identifier TID. With the concept of temporal layers, extracting a coded video sequence of lower temporal resolution from a given video sequence can be simply achieved by discarding all NAL units with TID larger than a selected value. As a result of such extracting the new CVS with another features is produced. In case of temporal scalability the main feature of new CVS is a picture rate (or FPS) which is reduced in respect to picture rate of parent CVS.

The specification of HEVC and VVC provide several signaling mechanisms for description of temporal dependencies between pictures such as:

* POC value with SPS parameters (sps\_max\_sub\_layers\_minus1, num\_units\_in\_tick, time\_scale …)
* Picture Timing SEI message
* Structure of Pictures SEI message
* Video Usability Information(VUI)

The parameters TargetLayerIdList and HighestTid could be set up by external means.

# Problem description

During transmission, storing or other manipulations with CVS with temporal sub-stream extraction the switching from one CVS with chosen Highest Temporal ID (HTID) to another CVS could not share the same HTID. At the time when HTID was defined according some external needs (like bandwidth adaptation or unequal error protection of temporal layers in error prone networks) the structure of the further CVS may have another temporal ID distribution as depicted on Figure 1.



Figure 1. Switching content with temporal sub-stream extraction

In the example provided on Figure 1 during transmission server decided to restrict Temporal Layer of SOP-1 downward to TID = 1 (from 60 FPS to 30 FPS). The current SOP0 have an extended GOP structure and if we continue to work within that restrictions (TID=1) then for current SOP0 the frame rate will be dropped down to 7.5 FPS. Such behavior is unexpected by user.

In order to follow the same temporal restriction the external application (network, MANE …) should perform some parsing and estimation:

* In case of SPS parameters: to get the Maximum Temporal ID (MTID) to parse picture rate corresponded to MTID and to find the TID corresponded to HTID of previous CVS. It is possible when fixed picture rate flags is equal to 1.
* In case of Picture Timing SEI message: The parsing of several messages is required because it is associated with each picture. Then to estimate picture rate according removal delays of each pictures.
* In case of SOP SEI message: The deriving and analysis of SOP is required in order to recognize appropriated TID with HTID of previous CVS.

Situation is further complicated by dynamically adaptation of SOPs TID distributions and picture timing parameters.

Most of the existing in specification mechanisms developed to support buffering, HRD, PLT and other constrains. This proposal try to provide simple possibility for upper to video abstraction layer applications to match requested picture rate with existing TID distribution. Such possibility is critical for video transmission and synchronization (audio, control, other effects like 4D/5D).

# Solution

Only encoder control the distribution of Temporal IDs and whether structure of pictures was modified during transmission and other adaptation the temporal layers could be controlled according predefined parameters. Two approaches are proposed below: by signaling in SEI message and by additional parameters in HRD design.

In a large majority of use case scenarios the distribution of temporal layer IDs among pictures is dyadic in order to support very efficient approach of bi-directional inter prediction with hierarchical structure of pictures (SOP). For such case the signaling the corresponded flag reflected dyadic temporal IDs distribution is proposed for both solutions.

## Specification Text changes in respect to HEVC

This document contains the draft text for changes to the High Efficiency Video Coding (HEVC) standard (Rec. ITU-T H.265 | ISO/IEC 23008-2) to specify additional supplemental enhancement information (SEI) messages for Sub-Layer Picture Rates.

*Replace D.2.1 with the following:*

**D.2.1 General SEI message syntax**

|  |  |
| --- | --- |
| sei\_payload( payloadType, payloadSize ) { | **Descriptor** |
| if( nal\_unit\_type  = =  PREFIX\_SEI\_NUT ) |  |
| if( payloadType  = =  0 ) |  |
| buffering\_period( payloadSize ) |  |
| else if( payloadType  = =  1 ) |  |
| pic\_timing( payloadSize ) |  |
| else if( payloadType  = =  2 ) |  |
| pan\_scan\_rect( payloadSize ) |  |
| else if( payloadType  = =  3 ) |  |
| … |  |
| else if( payloadType  = =  168 ) |  |
| frame\_field\_info( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  169 ) |  |
| pic\_rates( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  176 ) |  |
| three\_dimensional\_reference\_displays\_info( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  177 ) |  |
| depth\_representation\_info( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  178 ) |  |
| multiview\_scene\_info( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  179 ) |  |
| multiview\_acquisition\_info( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  180 ) |  |
| multiview\_view\_position( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  181 ) |  |
| alternative\_depth\_info( payloadSize ) /\* specified in Annex I \*/ |  |
| else if( payloadType  = =  200 ) |  |
| sei\_manifest( payloadSize ) |  |
| else if( payloadType  = =  201 ) |  |
| sei\_prefix\_indication( payloadSize ) |  |
| else |  |
| reserved\_sei\_message( payloadSize ) |  |
| else /\* nal\_unit\_type  = =  SUFFIX\_SEI\_NUT \*/ |  |
| if( payloadType  = =  3 ) |  |
| filler\_payload( payloadSize ) |  |
| else if( payloadType  = =  4 ) |  |
| user\_data\_registered\_itu\_t\_t35( payloadSize ) |  |
| else if( payloadType  = =  5 ) |  |
| user\_data\_unregistered( payloadSize ) |  |
| else if( payloadType  = =  17 ) |  |
| progressive\_refinement\_segment\_end( payloadSize ) |  |
| else if( payloadType  = =  22 ) |  |
| post\_filter\_hint( payloadSize ) |  |
| else if( payloadType  = =  132 ) |  |
| decoded\_picture\_hash( payloadSize ) |  |
| else if( payloadType  = =  146 ) |  |
| coded\_region\_completion( payloadSize ) |  |
| else |  |
| reserved\_sei\_message( payloadSize ) |  |
| if( more\_data\_in\_payload( ) ) { |  |
| if( payload\_extension\_present( ) ) |  |
| **reserved\_payload\_extension\_data** | u(v) |
| **payload\_bit\_equal\_to\_one** /\* equal to 1 \*/ | f(1) |
| while( !byte\_aligned( ) ) |  |
| **payload\_bit\_equal\_to\_zero** /\* equal to 0 \*/ | f(1) |
| } |  |
| } |  |

**F.14.2.12 Sub-Layer Picture Rates SEI message syntax**

|  |  |
| --- | --- |
| pic\_rates( payloadSize ) { | **Descriptor** |
| **pr\_max\_sub\_layers\_minus1** | u(3) |
| **pr\_time\_scale** | u(32) |
| **dyadic\_temporal\_nesting\_flag** | u(1) |
| for( i = dyadic\_temporal\_nesting\_flag\* pr\_max\_sub\_layers\_minus1; i <= pr\_max\_sub\_layers\_minus1; i++ ) { |  |
| **sub\_layer\_num\_units\_in\_tick[ i ]** | u(32) |
| } |  |
| } |  |

**F.14.3.12 Sub-Layer Picture Rates SEI message semantics**

**pr\_max\_sub\_layers\_minus1** plus 1 specifies the maximum number of temporal sub-layers that may be present in each CVS referring to the SPS. The value of pr\_max\_sub\_layers\_minus1 shall be in the range of 0 to 6, inclusive.

It is a requirement of bitstream conformance that the value of pr\_max\_sub\_layers\_minus1 in the picture rates SEI message is equal to the value of sps\_max\_sub\_layers\_minus1 in the SPS.

**pr\_time\_scale** is the number of time units that pass in one second. For example, a time coordinate system that measures time using a 27 MHz clock has a time\_scale of 27 000 000. The value of time\_scale shall be greater than 0.

It is a requirement of bitstream conformance that the value of pr\_time\_scale in the picture rates SEI message is equal to the value of vui time\_scale in the VUI if present.

**dyadic\_temporal\_nesting\_flag** indicates that the Temporal IDs are distributed hierarchically with dyadic dependencies. When not present the value of dyadic\_temporal\_nesting\_flag is inferred to be equal to 0.

**sub\_layer\_num\_units\_in\_tick**[ i ] is the number of time units of a clock operating at the frequency pr\_time\_scale Hz that corresponds to one increment (called a clock tick) of a clock tick counter. sub\_layer\_num\_units\_in\_tick[ i ] shall be greater than 0. A clock tick, in units of seconds, is equal to the quotient of sub\_layer\_num\_units\_in\_tick divided by pr\_time\_scale. For example, when the picture rate of a video signal is 25 Hz, pr\_time\_scale may be equal to 27 000 000 and sub\_layer\_num\_units\_in\_tick may be equal to 1 080 000, and consequently a clock tick may be equal to 0.04 seconds.

It is a requirement of bitstream conformance that the value of sub\_layer\_num\_units\_in\_tick[ pr\_max\_sub\_layers\_minus1 ] in the picture rates SEI message is equal to the value of vui\_num\_units\_in\_tick in the VUI if present.

When dyadic\_temporal\_nesting\_flag is equal to 0 the values of sub\_layer\_num\_units\_in\_tick[ i ] for i from 0 to pr\_max\_sub\_layers\_minus1-1 are derived as follows:

For each value of i in the range of 0 to pr\_max\_sub\_layers\_minus1-1, inclusive, the variable sub\_layer\_num\_units\_in\_tick[ i ] is derived as follows:

sub\_layer\_num\_units\_in\_tick[ i ] = (1)

sub\_layer\_num\_units\_in\_tick[ pr\_max\_sub\_layers\_minus1 ] / 2^(pr\_max\_sub\_layers\_minus1 – i - 1)

# Patent rights declaration(s)

**Huawei Technologies may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**