|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  14th Meeting: Vienna, Austria, AT, July 29 – Aug 2, 2013 | Document: JCTVC-N0236 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **HLS: Extensions of motion-constrained tile sets SEI message for interactivity** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Cheung Auyeung, Jun Xu 1730 N. First Street San Jose, CA 95112, USA | Tel: Email: | +1-408-352-4086 [Cheung.Auyeung@am.sony.com](mailto:Cheung.Auyeung@am.sony.com) [jun.xu@am.sony.com](mailto:jun.xu@am.sony.com) |
| *Source:* | Sony Electronics Inc. | | |

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

# Abstract

The motion-constrained tile sets in Range Extensions text specification: Draft 3 does not have the functionality to signal all tiles are independently decodable and to signal the regions of interest (ROI) that may have more than one tile per ROI. This contribution proposes an extension to the motion-constrained tile sets SEI message to provide this functionality. With this extension, it enables the functionality to re-define any independently decodable region of interest in a CVS without re-encoding the CVS at CTU level based on user interactivity. Consequently this proposal supports the interactivity needed for various applications such as interactive UHDTV application, dynamic high-quality zoom-in application, interactive on-demand, e-learning, smart surveillance, and etc.

# Introduction

In version 1 of the HEVC standard [1], when tiles\_enabled\_flag is equal to 1, each picture is constructed of rectangular regions known as tiles. In Range Extensions text specification: Draft 3, these tiles can be collected to form a collection of tile sets as regions of interest. These tile sets can be indicated as independently decodable by the motion-constrained tile sets SEI messages [2].

Such HEVC tile structure can be applied to interactive tiled streaming. An example of tiled streaming is an application where a multi-resolution panorama video can be split into videos which composed of different regions of interest in the panorama and with different spatial resolution [3, 4, 5]. A user can view a user dependent region of interest (ROI) and resolution. Tiled streaming use cases include interactive UHDTV application, dynamic high-quality zoom-in application, interactive on-demand e-learning, smart surveillance, and etc. These use cases of tiled streaming are currently being studied in the core experiments of DASH [6]. See Figure 1 for an example of tiled streaming application image.

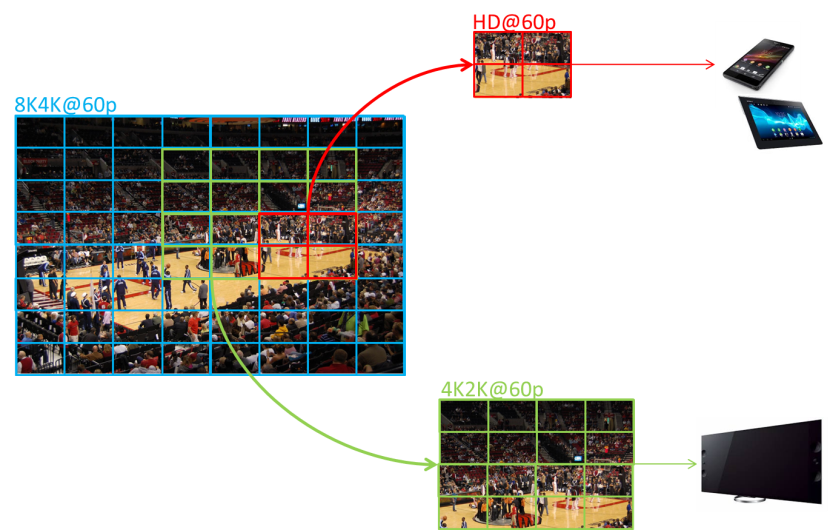


Figure Example of tiled streaming application

The characteristics of the interactive tiled streaming applications can be defined as follows:

* All tiles are independently decodable (motion-constrained tiles).
* The region of interests is not necessary determined at the beginning of the encode time of a CVS.
* Once the region of interests are encoded in a CVS, they *can* be modified based on user interactivity without re-encoding the CVS at the CTU level.

The current motion-constrained tile set SEI message is not appliable to interactive tiled streaming. If the current syntax is applied to interactive tiled streaming, each tile becomes one motion-constraied tile set with **exact\_sample\_value\_match\_flag**[ i ]=1. This has two issues:

* First, it losses the functionality to indicate the location of a region of interest with more than one tile.
* Second, the signaling become inefficient because each tile needed to signal the **exact\_sample\_value\_match\_flag** separately. In case of the example in Figure 1, it has 64 tile sets and 64 **exact\_sample\_value\_match\_flag** are signaled.

This proposal extends the syntax of a motion-constrained tile sets SEI message to provide improved signaling of motion constrained tile sets where all tiles are independently decodable so that the motion-constrained tile sets can be interactively re-defined for a CVS without re-encoding the CVS at the CTU level.

# Proposals

This contribution propose to add a flag to the motion-constrained tile sets SEI message to indicate that all tiles are independently decodable. The proposed syntax and semantics are highlighted in yellow as follows.

|  |  |
| --- | --- |
| motion\_constrained\_tile\_sets( payloadSize ) { | Descriptor |
| **all\_tiles\_exact\_sample\_value\_match\_\_flag** | u(1) |
| **num\_sets\_in\_message\_minus1** | ue(v) |
| for( i = 0; i <= num\_sets\_in\_message\_minus1; i++) { |  |
| **mcts\_id**[ i ] | ue(v) |
| **num\_tile\_rects\_in\_set\_minus1**[ i ] | ue(v) |
| for( j = 0; j <= num\_tile\_rects\_in\_set\_minus1[ i ]; j++) { |  |
| **top\_left\_tile\_index[**i **][**j **]** | ue(v) |
| **bottom\_right\_tile\_index**[ i ]**[**j **]** | ue(v) |
| } |  |
| if( ! **all\_tiles\_exact\_sample\_value\_match\_**\_flag) { |  |
| **exact\_sample\_value\_match\_flag**[ i ] | u(1) |
| } |  |
| } |  |
| } |  |

**all\_tiles\_exact\_sample\_value\_match\_flag** equal to 0 indicates that, within the coded video sequence, when the coding tree blocks that are outside any tile are not decoded and the boundaries of the tile is treated as picture boundaries for purposes of the decoding process, the value of each sample in the tile may not be exactly the same as the value of the same sample when all the coding tree blocks of the picture are decoded. all\_tiles\_exact\_sample\_value\_match\_flag equal to 1 indicates that, within the coded video sequence, when the coding tree blocks that do not belong to any tile are not decoded and the boundaries of the tile is treated as picture boundaries for purposes of the decoding process, the value of each sample in the tile would be exactly the same as the value of the sample that would be obtained when all the coding tree blocks of all pictures in the coded video sequence are decoded.

# Conclusion

When the motion-constrained tiles set in Range Extensions text specification Draft 3[2] is used to signal that all tiles are independently decodable, it losses the functionality to signal the locations of regions of interest with more than one tile. This contribution propose to add a flag to the motion-constrained tile sets SEI message to signal all tiles are independently decodable and maintains the capability to signal the locations of the regions of interest. This contribution also save signaling bits when all tiles are independently decodable.

# Reference

1. B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, and T. Wiegand (Eds.), “High efficiency video coding (HEVC) text specification draft 8”, JCTVC-L1003\_v34, 12th Meeting: Geneva, CH, 14–23 Jan. 2013.
2. D.Flynn, J.sole, T.Suzuki, “High Efficiency Video Coding (HEVC) Range Extensions text specification: Draft 3”, JCTVC-M1005, 13th Meeting: Incheon, KR, 18–26 Apr. 2013.
3. Franck Denoual, Hervé Le Floch, Frédéric Mazé, Eric Nassor, Nael Ouedraogo, Cyril Concolato, Jean Le Feuvre, “Interactive ROI streaming with DASH”, MPEG2013/M29232, Incheon, South Korea, April 2013.
4. Emmanuel Thomas, Rob Koenen, “Spatially segmented content description”, MPEG2013/m28883, Incheon, South Korea, April 2013.
5. Sally Hattori, Ohji Nakagami, Teruhiko Suzuki, “**H**LS: Extensions to Motion-constrained tile sets SEI message”, JCTVC-Nxxxx, 14th Meeting: Vienna, Austria, AT, July 29 – Aug 2, 2013.
6. A.Giladi, T.Stockhammer, EE coordinators, “Descriptions of Core Experiments on DASH Amendement”, N13499, MPEG2013, Incheon, Korea, April 2013.

# Patent rights declaration(s)

**Sony Corp. and Sony Electronics Inc. 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).**