|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  31st Meeting: San Diego, US, 13–20 Apr. 2018 | Document: JCTVC-AE0025-v2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **On lightweight bitstream merging** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | **Ye-Kui Wang** Huawei Technologies Co., Ltd. Telesis Court, San Diego CA 92121, USA  **Huameng Fang Yuqun Fan** 4#24C TiXiangMingYuan XiangMei Road, Futian district, Shenzhen, Guangdong, China  **Jianle Chen** Huawei Technologies Co., Ltd. 2330 Central Expy, Santa Clara CA 95050, USA | Tel: Email: | +1-858-754-3673 [yekui.wang@huawei.com](mailto:yekui.wang@huawei.com)  +86-755-2897-9770 [fanghuameng@huawei.com](mailto:fanghuameng@huawei.com)  +86-755-2897-5725 [fanyuqun@huawei.com](mailto:fanyuqun@huawei.com)  +1-408-330-5334 [jianle.chen@huawei.com](mailto:jianle.chen@huawei.com) |
| *Source:* | Huawei Technologies Co., Ltd. | | |

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

# Abstract

This contribution proposes a new HEVC SEI message to indicate whether the bitstream is constrained in a manner such that the bitstream can be lightweight merged/rewritten with other similarly-constrained bitstreams into one conforming bitstream without unexpected artifacts. This contribution defines lightweight merging/rewriting of multiple bitstreams into one bitstream as a process that generates one HEVC bitstream out of multiple HEVC bitstreams without changing block-level coding results, similarly as the motion-constrained tile set (MCTS) sub-bitstream extraction process that is specified as part of the semantics of the MCTSs extraction information sets SEI message. It is asserted that such lightweight merging/rewriting of multiple HEVC bitstreams into one HEVC bitstream is the basis of the HEVC-based viewport-dependent omnidirectional media format (OMAF) video profile that is specified in version 1 of OMAF, a media format standard that has been recently specified by MPEG for enabling of omnidirectional media applications, focusing on 360° video, images, and audio, as well as associated timed text.

# Introduction

Omnidirectional media format (OMAF), part 2 of ISO/IEC 23090, is a media format standard that has been recently specified by MPEG for enabling of omnidirectional media applications, focusing on 360° video, images, and audio, as well as associated timed text.

The HEVC-based viewport-dependent OMAF video profile is a media profile specified in OMAF. The key feature of this OMAF media profile is that it supports viewport-dependent streaming of omnidirectional video by utilizing a file format feature for lightweight merging/rewriting of multiple HEVC bitstreams into one HEVC bitstream, wherein one HEVC bitstream is generated out of multiple HEVC bitstreams without changing block-level coding results, similarly as the motion-constrained tile set (MCTS) sub-bitstream extraction process that is specified as part of the semantics of the MCTSs extraction information sets SEI message. The output bitstream usually consists of multiple MCTSs.

In lightweight merging/rewriting of multiple HEVC bitstreams into one HEVC bitstream, when some motion vectors of some pictures in an input bitstream point to sample positions out of the picture boundary, the output bitstream can still be a conforming HEVC bitstream that satisfy all the specified bitstream constraints, but the decoded picture can have unexpected, serious mismatch.

Moreover, when decoding an MCTS of the output bitstream, which is a conforming HEVC bitstream, it is still possible that the usage of motion vector candidates for temporal motion vector prediction are derived from blocks outside the MCTS, and when this occurs, again the decoded picture can have unexpected, serious mismatch artifacts.

Therefore, to be able to be lightweight merged/rewritten with other bitstreams into one conforming bitstream without unexpected artifacts, the input bitstreams need to be constrained such that the above issues do not occur. Actually, the constraints needed are similarly as the MCTS constraints that are specified as part of the semantics of the temporal MCTSs SEI message, which has been amended in JCTVC-AC1005-v2, as follows:

...

The temporal motion-constrained tile sets SEI message indicates that the following constraints apply:

– No sample values outside each identified tile set or outside the picture are referenced for inter prediction.

– For PUs located directly left of the right tile boundary of each identified tile set except the last one at the bottom right, the following applies when CuPredMode[ xPb ][ yPb ] is equal to MODE\_INTER, where ( xPb, yPb ) specifies the top-left sample of the corresponding luma prediction block relative to the top-left sample of the current picture:

– With the number of spatial merging candidates numSpatialMergeCand derived as follows:

numSpatialMergeCand = availableFlagA0 + availableFlagA1 +  (D‑42)  
 availableFlagB0 + availableFlagB1 + availableFlagB2

where availableFlagA0, availableFlagA1, availableFlagB0, availableFlagB1, and availableFlagB2 are the output of the derivation process for spatial merging candidates specified in clause 8.5.3.2.3, the following applies:

– If numSpatialMergeCand is equal to 0, merge\_flag[ xPb ][ yPb ] is equal to 0.

– Otherwise (numSpatialMergeCand is greater than 0), merge\_idx[ xPb ][ yPb ] is in the range of 0 to numSpatialMergeCand − 1, inclusive.

– With the number of spatial motion vector predictor candidates numSpatialMvpCand derived as follows:

if ( availableFlagLXA )  
 numSpatialMvpCand = availableFlagLXA + ( ( mvLXA  !=  mvLXB ) ? availableFlagLXB : 0 )  
else (D‑43)  
 numSpatialMvpCand = availableFlagLXB

where availableFlagLXA, availableFlagLXB, mvLXA, and mvLXB are the output of the derivation process for motion vector predictor candidates from neighbouring prediction unit partitions specified in clause 8.5.3.2.7, the following applies:

– If numSpatialMvpCand is equal to 0, mvp\_l0\_flag[ xPb ][ yPb ] and mvp\_l1\_flag[ xPb ][ yPb ] is equal to 1.

– Otherwise (numSpatialMvpCand is greater than 0), mvp\_l0\_flag[ xPb ][ yPb ] and mvp\_l1\_flag[ xPb ][ yPb ] is in the range of 0 to numSpatialMvpCand − 1, inclusive.

NOTE 1 – The first constraint restricts motion vectors to point to full-sample locations inside each identified tile set and to fractional-sample locations that require only full-sample locations inside each identified tile set for interpolation. The second constraint restricts the usage of motion vector candidates derived from blocks outside each identified tile set.

*...*

Currently, given an HEVC bitstream, whether it satisfies these constraints can only be concluded by parsing into the bitstream to block-level details such as motion information and performing the derivation process for motion vector components and reference indices, possibly for all the pictures.

Therefore, to be able to easily indicate whether these constraints are satisfied for an HEVC bitstream, a new SEI message is proposed.

# Proposal

The proposed new SEI message is named lightweight merging indication SEI message. The syntax and semantics of this SEI message are as follows.

|  |  |
| --- | --- |
| lightweight\_merging\_indication( payloadSize ) { | **Descriptor** |
| } |  |

When a lightweight merging indication SEI message is present for any picture of a CLVS, there shall be a lightweight merging indication SEI message present for the first picture of the CLVS.

The presence of the lightweight merging indication SEI message in a CLVS indicates that the following constraints apply for each picture of the CLVS:

– There shall be no sample values outside the picture that are referenced for inter prediction.

– For PUs located on the right-side picture boundary except the last one at the bottom-right corner of the picture, the following applies when CuPredMode[ xPb ][ yPb ] is equal to MODE\_INTER, where ( xPb, yPb ) specifies the top-left sample of the corresponding luma prediction block relative to the top-left sample of the current picture:

– With the number of spatial merging candidates numSpatialMergeCand derived as follows:

numSpatialMergeCand = availableFlagA0 + availableFlagA1 + (D‑X)  
 availableFlagB0 + availableFlagB1 + availableFlagB2

where availableFlagA0, availableFlagA1, availableFlagB0, availableFlagB1, and availableFlagB2 are the output of the derivation process for spatial merging candidates specified in clause 8.5.3.2.3, the following applies:

– If numSpatialMergeCand is equal to 0, merge\_flag[ xPb ][ yPb ] shall be equal to 0.

– Otherwise (numSpatialMergeCand is greater than 0), merge\_idx[ xPb ][ yPb ] shall be in the range of 0 to numSpatialMergeCand − 1, inclusive.

– With the number of spatial motion vector predictor candidates numSpatialMvpCand derived as follows:

if ( availableFlagLXA )  
 numSpatialMvpCand = availableFlagLXA +  (D‑X)  
 ( ( mvLXA  !=  mvLXB ) ? availableFlagLXB : 0 )  
else  
 numSpatialMvpCand = availableFlagLXB

where availableFlagLXA, availableFlagLXB, mvLXA, and mvLXB are the output of the derivation process for motion vector predictor candidates from neighbouring prediction unit partitions specified in clause 8.5.3.2.7, the following applies:

– If numSpatialMvpCand is equal to 0, mvp\_l0\_flag[ xPb ][ yPb ] and mvp\_l1\_flag[ xPb ][ yPb ] shall both be equal to 1.

– Otherwise (numSpatialMvpCand is greater than 0), mvp\_l0\_flag[ xPb ][ yPb ] and mvp\_l1\_flag[ xPb ][ yPb ] shall both be in the range of 0 to numSpatialMvpCand − 1, inclusive.

NOTE – The first constraint restricts that motion vectors point either to full-sample locations inside the picture or to fractional-sample locations that require only full-sample locations inside the picture for interpolation. The second constraint restricts that, when the bitstream is merged with other similarly-constrained bitstreams into one conforming bitstream in a fashion where block-level coding results remain unchanged, e.g., similarly as the MCTS sub-bitstream extraction specified as part of the semantics of the MCTSs extraction information sets SEI message, in decoding of the entire merged bitstream, for blocks of the "sub-picture" corresponding to this bitstream, there won't be motion vector candidates for temporal motion vector prediction derived from blocks outside the "sub-picture".

# Patent rights declaration(s)

**Huawei Technologies Co., Ltd. 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).**