|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  12th Meeting: Geneva, CH, 14–23 Jan. 2013 | Document: JCTVC-L0049 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **SEI message: independently decodable regions based on tiles** | | |
| *Status:* | Input document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Yan Ye, Yong He, Yuwen He 9710 Scranton Rd, Suite 250 San Diego, CA 92121  Xiaofeng Yang, Peiyu Yue, Yuanyuan Zhang Nanjing R&D Center, 101 Software Avenue, Yuhuatai District, Nanjing, P.R.China 200012  Michael Horowitz  1450-1055 W. Hastings Street  Vancouver, BC V6E 2E9  Canada | Tel:  Email:  Tel:  Email:  Tel:  Email: | +1 858.210.4803 [yan.ye@interdigital.com](mailto:yan.ye@interdigital.com)  +86 25 56620730 [yx.yangxiaofeng@huawei.com](mailto:yx.yangxiaofeng@huawei.com)  +1 778.331.3432 x323 [michael@ebriskvideo.com](mailto:michael@ebriskvideo.com) |
| *Source:* | InterDigital Communications, LLC, Huawei Technologies Co., Ltd., and eBrisk Video | | |

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

# Abstract

This is a follow-up proposal to JCTVC-K0116 and JCTVC-K0248. At the 11th JCT-VC meeting, JCTVC-K0116 and JCTVC-K0248 proposed an SEI message to support independently decodable regions using tiles. In this contribution, simulation results are provided to show that, for typical use case scenarios, the restrictions on motion compensated prediction as required by the proposed SEI message incur RD performance penalty of 1.4% and 2.5% when loop filter across tile boundary is disabled and enabled, respectively. Subjective viewing of the reconstructed video did not reveal noticeable coding artefacts.

# Introduction

A number of applications can benefit from the capability to independently decode a portion of the video frames from a video sequence. For example, in frame compatible 3D video (Figure 1), the user may have 2D-only display. In this case, the decoder will need to only decode half of the frame, e.g., the left view. In multi-way video conference system (Figure 2), when a speaker is giving a presentation and/or sharing a PPT, it is often useful to be able to display only a portion of the video screen that contains the presentation and/or the speaker on user’s demand. Another example is the e-learning system. Recently, Stanford researchers applied independently decodable region concept to H.264/AVC to an online e-learning project called ClassX [2]. Figure 3 shows an example of ClassX Mobile with user zooming into the video to focus on the content on the blackboard. As the recent boom of mobile video applications continues, providing support for independently decodable regions in HEVC is particularly crucial, as it allows the decoder the option to significantly reduce decoding complexity by skipping decoding and displaying of unwanted regions in the video frame.

The HEVC Main profile supports tiles, which are independently decodable bitstream units that represent rectangular regions of a video frame. This makes tiles a natural choice to support independently decodable regions in HEVC. At the last meeting, JCTVC-K0116 and JCTVC-K0248 proposed tiles-based SEI message for this purpose.

 

Fig. 1 Frame compatible 3D video



Fig. 2 Multi-point video conference



Fig. 3 ClassX Mobile

# Tile sections

In general, assume the independently decodable region has overlap with N tiles. These N tiles together form a so called tile section. For example, in Fig. 4, the two tiles on the left form a tile section.

In current HEVC draft, in-loop filtering can be applied across tile boundaries if loop\_filter\_across\_tiles\_enabled\_flag (in PPS) is set to 1. If the decoder chooses to decode only the tile section, but not the remaining tiles outside of the tile section, then it will not have neighbouring samples from tiles outside the current tile section to perform proper in-loop filtering, including SAO and deblocking. To avoid temporal error propagation, the samples that are affected by non-conformant decoding behaviour should not be used by MCP to predict samples in future pictures. That is, MCP should not only be constrained to be within the tile section, a narrow range of sample locations along the border of the tile section should be excluded from MCP as well. As shown in Figure 4, a band, 4 samples wide along the right border of the tile section (shaded area), should be excluded from MCP. The width of the 4-sample-wide band is determined by the length of the deblocking filter, which has larger neighbourhood dependency than SAO. Although decoding only samples in the tile section causes samples within the narrow band (shaded area) to have values that deviate from their correctly decoded values, such deviation will not cause temporal error propagation due to the MCP restriction thus imposed.

In Figure 4, the tile section only has one boundary (out of four boundaries) that is not a picture boundary. In general, a tile section can have more than one boundary that is not a picture boundary but is a boundary with other tiles. In the general case, any boundary of a tile section that is not a picture boundary should have the narrow band of 4 samples wide (shaded area in Figure 4); samples within these narrows bands should not be used for MCP.



Fig. 4 Constraint on range of MCP to support tile-based independently decodable regions

When loop\_filter\_across\_tiles\_enabled\_flag is set to 0, the size of the narrow band for which MCP restriction is applied is set to 0. This can be useful for applications such as frame compatible 3D video (Figure 1) and multi-point video conference (Figure 2), for which discontinuity typically already exists inside the video frame.

# Proposed SEI message

## Proposed SEI message syntax

|  |  |
| --- | --- |
| tile\_section ( payloadSize ) { | Descriptor |
| **num\_tile\_sections** | ue(v) |
| for( i = 0; i < num\_tile\_sections; i++ ) { |  |
| **num\_tiles\_in\_section\_minus1**[ i ] | ue(v) |
| for( j = 0; j <= num\_tiles\_in\_section\_minus1[ i ]; j++ ) |  |
| **tile\_index**[ i ][ j ] | ue(v) |
| **tile\_section\_exact\_match\_flag** [ i ] | u(1) |
| } |  |
| } |  |

## Proposed SEI message semantics

The tile section SEI message indicates to the decoder which tiles in the picture form independently decodable regions. A tile section defines an area within a video picture for which inter prediction is constrained. Inter prediction of any samples within a tile section shall use no samples from outside the tile section. When loop filter is applied across tile section boundaries, inter prediction of any samples within a tile section should additionally not depend on any samples within 4 samples away from the tile section boundary, unless the tile section boundary is also a picture boundary.

A tile section SEI message persists for all pictures in decoding order, from the coded picture to which the SEI message is associated (inclusive), to the coded picture to which the next tile section SEI message in decoding order is assoicated (exclusive), or to the last access unit in the coded video sequence (inclusive).

**num\_tile\_sections** specifies the number of tile sections in the picture. num\_ tile\_sections shall be in the range of 0 to (num\_tile\_columns\_minus1 + 1) \* (num\_tile\_rows\_minus1 + 1), inclusive.

**num\_tiles\_in\_ section\_minus1**[ i ] plus 1 specifies the number of tiles in the i-th tile section. num\_tiles\_in\_ section\_minus1[ i ] shall be in the range of 0 to (num\_tile\_columns\_minus1 + 1) \* (num\_tile\_rows\_minus1 + 1) − 1, inclusive.

**tile\_index**[ i ] [ j ] specifies the index of the j-th tile in the i-th tile section in the picture. tile\_index[ i ] [ j ] shall be in the range of 0 to (num\_tile\_columns\_minus1 + 1) \* (num\_tile\_rows\_minus1 + 1) − 1, inclusive. The index values of the tiles in a picture are assigned by starting from 0 in the top left corner, and incrementing by 1 following the raster scan order.

**tile\_section\_exact\_match\_flag**[ i ] indicates whether decoding only the i-th tile section produces reconstructed samples that are exactly the same as decoding the entire picture.

# Simulation results

At the 11th JCT-VC meeting in Shanghai, the concept of tiles-based independently decodable regions, as proposed in JCTVC-K0116 and JCTVC-K0248, received non-proponent support. The remaining concerns about the proposals were possible visual artefacts and potential coding efficiency loss due to the restriction placed on motion compensated prediction. In this contribution, the proposed MCP restriction was implementation on HM8.1. The video frame is partitioned into two horizontal, even tiles (--NumTileColumnsMinus1=1 --UniformSpacingIdc=1). MCP is constrained to only use samples within the region covered by each tile, respectively. If loop filter is enabled across tile boundary (--LFCrossTileBoundaryFlag=1), MCP is further constrained to exclude samples that are no more than 4 samples away from the tile boundary in the middle of the frame.

Only Class A and B sequences were used in the simulation, since the concept of independently decodable regions is less applicable to frame sizes smaller than 1080p. QP 22 was used. Other test conditions were kept the same as the common test conditions. Table 1 shows the performance impact by imposing the MCP restriction for LFCrossTileBoundaryFlag = 0 (0-sample band). The bit rate is increased by 1.4% on average, with minimal PSNR difference. Table 2 shows the performance impact by imposing the MCP restriction for LFCrossTileBoundary = 1 (4-sample band). The performance loss is slightly larger than Table 1 (which is expected as the MCP restriction for this case is slightly more severe); on average the bit rate is increased by 2.5%, and PSNR loss is negligible.

The reconstructed video was viewed at normal frame rate. No obvious coding artefacts were observed.

Table . Performance impact with MCP restriction when LFCrossTileBoundaryFlag = 0 (0-sample band), average rate increase is 1.4%

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | HM8.1 | | | | HM8.1+MCP restriction | | | | % rate increase | Y PSNR drop |
| kbps | Y psnr | U psnr | V psnr | kbps | Y psnr | U psnr | V psnr |
| Class A RA | Traffic | 13157.84 | 41.66 | 41.54 | 44.22 | 13328.33 | 41.65 | 41.54 | 44.21 | 1.3% | -0.01 |
| PeopleOnStreet | 32763.69 | 40.18 | 44.89 | 44.67 | 33036.22 | 40.17 | 44.87 | 44.66 | 0.8% | -0.01 |
| Nebuta | 216482.93 | 39.04 | 39.14 | 37.71 | 217066.48 | 39.04 | 39.13 | 37.70 | 0.3% | -0.01 |
| SteamLocomotive | 23610.19 | 41.37 | 46.22 | 45.89 | 23795.98 | 41.37 | 46.21 | 45.89 | 0.8% | -0.01 |
| Class B RA | Kimono | 4792.63 | 41.61 | 43.47 | 45.26 | 4855.91 | 41.59 | 43.46 | 45.25 | 1.3% | -0.01 |
| ParkScene | 7663.60 | 40.07 | 42.38 | 43.79 | 7769.78 | 40.06 | 42.38 | 43.78 | 1.4% | -0.01 |
| Cactus | 18278.54 | 38.48 | 40.04 | 43.60 | 18506.47 | 38.47 | 40.04 | 43.59 | 1.2% | -0.01 |
| BasketballDrive | 17362.12 | 39.16 | 43.77 | 44.99 | 17624.34 | 39.15 | 43.77 | 44.97 | 1.5% | -0.01 |
| BQTerrace | 39673.70 | 37.42 | 42.11 | 44.29 | 40139.15 | 37.41 | 42.11 | 44.29 | 1.2% | -0.01 |
| Class B LDB | Kimono | 5209.49 | 41.68 | 43.26 | 44.75 | 5324.40 | 41.66 | 43.26 | 44.75 | 2.2% | -0.02 |
| ParkScene | 7947.19 | 39.91 | 41.93 | 43.07 | 8112.59 | 39.89 | 41.92 | 43.06 | 2.1% | -0.01 |
| Cactus | 19903.27 | 38.69 | 40.11 | 43.36 | 20184.61 | 38.68 | 40.10 | 43.35 | 1.4% | -0.01 |
| BasketballDrive | 19815.99 | 39.42 | 43.64 | 44.84 | 20252.70 | 39.40 | 43.62 | 44.82 | 2.2% | -0.01 |
| BQTerrace | 52655.31 | 38.23 | 41.95 | 44.08 | 53390.23 | 38.23 | 41.93 | 44.07 | 1.4% | -0.01 |

Table 2. Performance impact with MCP restriction when LFCrossTileBoundaryFlag = 1 (4-sample band), average rate increase is 2.5%

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | HM8.1 | | | | HM8.1+MCP restriction | | | | % rate increase | Y PSNR drop |
| kbps | Y psnr | U psnr | V psnr | kbps | Y psnr | U psnr | V psnr |
| Class A RA | Traffic | 13155.03 | 41.66 | 41.54 | 44.21 | 13513.16 | 41.64 | 41.53 | 44.20 | 2.7% | -0.02 |
| PeopleOnStreet | 32765.69 | 40.17 | 44.88 | 44.66 | 33267.28 | 40.17 | 44.85 | 44.64 | 1.5% | -0.01 |
| Nebuta | 216469.62 | 39.04 | 39.13 | 37.70 | 217172.29 | 39.04 | 39.13 | 37.69 | 0.3% | -0.01 |
| SteamLocomotive | 23626.29 | 41.37 | 46.21 | 45.89 | 23836.37 | 41.36 | 46.21 | 45.89 | 0.9% | -0.01 |
| Class B RA | Kimono | 4791.76 | 41.60 | 43.46 | 45.26 | 4900.66 | 41.59 | 43.45 | 45.24 | 2.3% | -0.02 |
| ParkScene | 7659.24 | 40.07 | 42.38 | 43.78 | 7881.50 | 40.04 | 42.36 | 43.77 | 2.9% | -0.03 |
| Cactus | 18274.74 | 38.48 | 40.03 | 43.59 | 19104.61 | 38.46 | 40.02 | 43.56 | 4.5% | -0.02 |
| BasketballDrive | 17364.11 | 39.16 | 43.77 | 44.98 | 17687.30 | 39.14 | 43.75 | 44.96 | 1.9% | -0.01 |
| BQTerrace | 39673.09 | 37.42 | 42.10 | 44.28 | 40411.13 | 37.41 | 42.10 | 44.28 | 1.9% | -0.01 |
| Class B LDB | Kimono | 5207.92 | 41.68 | 43.25 | 44.75 | 5374.80 | 41.66 | 43.25 | 44.74 | 3.2% | -0.02 |
| ParkScene | 7948.53 | 39.90 | 41.92 | 43.07 | 8222.70 | 39.89 | 41.91 | 43.06 | 3.4% | -0.02 |
| Cactus | 19918.55 | 38.69 | 40.09 | 43.35 | 20919.17 | 38.68 | 40.08 | 43.33 | 5.0% | -0.01 |
| BasketballDrive | 19815.24 | 39.41 | 43.63 | 44.83 | 20339.15 | 39.40 | 43.61 | 44.81 | 2.6% | -0.01 |
| BQTerrace | 52660.01 | 38.23 | 41.94 | 44.08 | 53641.76 | 38.22 | 41.93 | 44.06 | 1.9% | -0.01 |

Additional test for the frame compatible 3D use case is also conducted. The 3DV test sequence Poznan\_Hall2 [3] is converted into frame compatible 3D format using the software tool proposed in [4]. One frame is shown in Fig. 4 as an example.



Fig. 4 Poznan\_Hall2 sequence in side-by-side frame compatible format

The R-D performance of the frame compatible Poznan\_Hall2 sequence is provided in Table 3. Random Access with QP = 32 and LFCrossTileBoundaryFlag = 0 is used. The reconstructed video was viewed at normal frame rate. No visible coding artefacts were observed. Fig. 4 shows one frame of the reconstructed frame compatible Poznan\_Hall2 sequence.

Table 3. Performance impact with MCP restriction for MFC sequence (RA, QP=32, LF=0)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | HM8.1 | | | | HM8.1+MCP restriction | | | | % rate increase | Y PSNR drop |
| kbps | Y psnr | U psnr | V psnr | kbps | Y psnr | U psnr | V psnr |
| Poznan-Hall2 | 341.34 | 39.71 | 45.35 | 45.17 | 344.82 | 39.71 | 45.34 | 45.18 | 1.0% | 0.00 |



Fig. 4 One reconstructed picture of Poznan\_Hall2 in frame compatible format

# Conclusion

This contribution proposes an SEI message that allows HEVC to support independently decodable regions using tiles. Simulation results were presented to show that the MCP restriction required by the proposed SEI message does not cause significant performance penalty or visible coding artefacts. It is therefore suggested to adopt this SEI message into HEVC WD.

# Patent rights declaration(s)

**InterDigital Communications, LLC may have IPR 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).**

**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).**

**eBrisk Video 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).**

# References

1. B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, T. Wiegand, High efficiency video coding (HEVC) text specification draft 9, Document no. JCTVC-K1003, October 2012.
2. A. Mavlankar, B. Girod, “Spatial-Random-Access-Enabled Video Coding for Interactive Virtual Pan/Tilt/Zoom Functionality,” IEEE Transactions on Circuits and Systems for Video Technology. vol. 21, no. 5, pp. 577-588, May 2011.
3. MPEG Video and Requirement, Call for Proposals on 3D Video Coding Technology, MPEG2011/N12036, March 2011
4. A. Leontaris, P. V. Pahalawatta, Y. He, A. M. Tourapis, and W. J. Husak, “Reformatting tools for frame-compatible and full-resolution stereoscopic content,” m20111, Geneva, Switzerland, March 2011.