|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  11th Meeting: Shanghai, CN, 10–19 Oct. 2012 | Document: JCTVC-K0248 r2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **ROI tile sections** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Yan Ye, Yuwen He, Yong He 9710 Scranton Rd, Suite 250 San Diego, CA 92121  Michael Horowitz  1450-1055 W. Hastings Street  Vancouver, BC V6E 2E9  Canada | Tel:  Email:  Tel:  Email: | +1 858.210.4803 [yan.ye@interdigital.com](mailto:yan.ye@interdigital.com)  +1 778.331.3432 x323 [michael@ebriskvideo.com](mailto:michael@ebriskvideo.com) |
| *Source:* | InterDigital Communications and eBrisk Video | | |

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

# Abstract

HEVC draft specification supports tiles in the Main profile, which can be used to provide high-level parallelization of encoding and decoding. Though the current design does not support it, tiles can also be used for region of interest coding. This contribution proposes to add syntax support to enable region of interest coding based on tiles.

# Motivations

The HEVC draft specification [1] supports tiles that may be used for high-level parallelization of encoding and decoding. Though the current design does not support it, tiles can also be used to provide region of interest coding. Region of interest coding is a useful feature for many video applications. For example, when high resolution video is displayed on a mobile device with limited screen size, instead of the entire picture, the user often chooses to view only a portion of the picture by panning, zooming, and so on. Stanford researchers recently applied Region of Interest coding for e-learning in a project called ClassX [2]. Figure 1 shows an example of ClassX Mobile with user zooming into the video. The underlying coding technology for ClassX is based on H.264/AVC. It follows that there exist applications for which ROI feature support in HEVC would be beneficial.



Figure 1. ClassX Mobile



Figure . Region of Interest and tiles

Figure 2shows an example where the picture is partitioned into 4 tiles, and where the ROI covers only the 2 tiles on the left. In the current HEVC draft spec, tiles 0 through 3 can be independently decoded for any given picture. Although decoding only the 2 tiles on the left is sufficient for displaying the ROI of the current picture, all tiles in the current picture still must be decoded such that future pictures can be correctly decoded using motion compensated prediction, that is, unless the current picture is a non-reference picture.

In order to improve support for ROI coding using HEVC, it is proposed to constrain the range of motion compensated prediction to tiles, which are already independently decode-able units within one picture, and to signal such constraint in the bitstream to the decoder. We next discuss the necessary constraint on the range of MCP that the encoder must enforce such that decoding error is guaranteed to be localized at the decoder should it choose to only decode the tiles in which the ROI is contained.

# ROI tile sections

In general, assume the ROI has overlap with N tiles. These N tiles together form a so called ROI tile section. For example, in Figure 2, the two tiles on the left form an ROI tile section. In current HEVC draft, in-loop filtering can be applied across tile boundaries if loop\_filter\_across\_tiles\_enabled\_flag is set to 1. If the decoder chooses to decode only the ROI tile section, but not the remaining non-ROI tiles, then it will not have neighbouring samples from tiles outside the ROI 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 ROI tile section, a narrow range of sample locations along the border of the ROI tile section should be excluded from MCP as well. As shown in Figure 3, a band, 4 samples wide along the right border of the ROI 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. Note that decoding only samples in the ROI tile section causes samples within the narrow band (shaded area) to have values that in general deviate from their correctly decoded values.

In Figure 3, the ROI tile section only has one boundary (out of four boundaries) that is not a picture boundary. In general, an ROI tile section can have more than one boundary that is not a picture boundary but is a boundary with other non-ROI tiles. In the general case, any boundary of an ROI tile section that is not a picture boundary should have the narrow band of 4 samples wide (shaded area in Figure 3); samples within these narrows bands should not be used for MCP.



Figure . Constraint on range of MCP to enable ROI only decoding

Table 1 shows syntax for ROI tile sections.

Table 1. ROI tile sections

|  |  |
| --- | --- |
| ROI\_tile\_section () { | Descriptor |
| **num\_ROI\_tile\_sections** | ue(v) |
| if(num\_ROI\_tile\_sections > 0) |  |
| for( i = 0; i < num\_ROI\_tile\_sections; i++ ) { |  |
| **num\_tiles\_in\_ROI\_section\_minus1**[ i ] | u(v) |
| for( j = 0; j <= num\_tiles\_in\_ROI\_section\_minus1[ i ]; j++ ) |  |
| **tile\_index**[ i ][ j ] | u(v) |
| } |  |
| } |  |

**num\_ROI\_tile\_sections** specifies the number of ROI tile sections in the picture. num\_ROI\_tile\_sections shall be in the range of 0 to (num\_tile\_columns\_minus1 + 1) \* (num\_tile\_rows\_minus1 + 1), inclusive. If num\_ROI\_tile\_sections is not present, it is inferred to be 0.

NOTE – an ROI tile section defines an area within a video picture for which motion compensated prediction is constrained. Motion compensated prediction of any samples within an ROI tile section should not depend on any samples from outside the ROI tile section, or from no more than 4 samples away from any ROI tile section boundary, unless the ROI tile section boundary is also a picture boundary. s

**num\_tiles\_in\_ROI\_section\_minus1**[ i ] plus 1 specifies the number of tiles in the i-th ROI tile section. num\_tiles\_in\_ROI\_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 ROI 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.

NOTE – When num\_ROI\_tile\_section is equal to 0, no ROI tile section is defined for the pictures in the video sequence. That is, there is no constraint on the range of motion compensated prediction for all samples in the picture. When num\_ROI\_tile\_section is equal to or greater than 1, composition of each of the ROI tile sections is signaled using tile\_index[ i ][ j ].

As discussed above, the ROI tile section allows a decoder to decode only the ROI tile section (and to display only the ROI) without temporal error propagation; at any time instance, all samples will be correctly decoded except for a 4 sample wide band along some of the ROI tile section boundaries. We consider the VUI to be the most suitable place to carry the ROI tile section syntax. We also provide an alternative option that signals the ROI tile sections using SEI message.

## An example ROI tile section

An encoder can partition the video picture into tiles according to the position and the dimension of the ROI. As shown in Figure 4, the encoder partitions the picture into 9 tiles, with the center tile T4 (blue shaded area) largely overlapping with the ROI (orange area), and a small margin is left between the tile T4 and the ROI to accommodate for the case when loop\_filter\_across\_tiles\_enabled\_flag is set to 1. To enable ROI-only decoding, the ROI tile section is defined to contain only T4; that is, in Table 1, num\_ROI\_tile\_sections = 1, num\_tiles\_in\_ROI\_section\_minus1[0] = 0, and tile\_index[0][0] = 4. If and when the ROI moves to another position in the picture at a later time instance, then the tile partition parameters can be changed accordingly to align with the ROI’s new position.



Figure 4. Example ROI tile section

# Proposed text change of option 1: using VUI to signal ROI tile sections

|  |  |
| --- | --- |
| vui\_parameters( ) { | Descriptor |
| **aspect\_ratio\_info\_present\_flag** | u(1) |
| if( aspect\_ratio\_info\_present\_flag ) { |  |
| **aspect\_ratio\_idc** | u(8) |
| if( aspect\_ratio\_idc = = Extended\_SAR ) { |  |
| **sar\_width** | u(16) |
| **sar\_height** | u(16) |
| } |  |
| } |  |
| **overscan\_info\_present\_flag** | u(1) |
| if( overscan\_info\_present\_flag ) |  |
| **overscan\_appropriate\_flag** | u(1) |
| **video\_signal\_type\_present\_flag** | u(1) |
| if( video\_signal\_type\_present\_flag ) { |  |
| **video\_format** | u(3) |
| **video\_full\_range\_flag** | u(1) |
| **colour\_description\_present\_flag** | u(1) |
| if( colour\_description\_present\_flag ) { |  |
| **colour\_primaries** | u(8) |
| **transfer\_characteristics** | u(8) |
| **matrix\_coefficients** | u(8) |
| } |  |
| } |  |
| **chroma\_loc\_info\_present\_flag** | u(1) |
| if( chroma\_loc\_info\_present\_flag ) { |  |
| **chroma\_sample\_loc\_type\_top\_field** | ue(v) |
| **chroma\_sample\_loc\_type\_bottom\_field** | ue(v) |
| } |  |
| **neutral\_chroma\_indication\_flag** | u(1) |
| **field\_seq\_flag** | u(1) |
| **hrd\_parameters\_present\_flag** | u(1) |
| if( hrd\_parameters\_present\_flag ) |  |
| hrd\_parameters( 1, sps\_max\_sub\_layers\_minus1 ) |  |
| **bitstream\_restriction\_flag** | u(1) |
| if( bitstream\_restriction\_flag ) { |  |
| **tiles\_fixed\_structure\_flag** | u(1) |
| **num\_ROI\_tile\_sections** | u(v) |
| if(num\_ROI\_tile\_sections > 0) |  |
| for( i = 0; i < num\_ROI\_tile\_sections; i++ ) { |  |
| **num\_tiles\_in\_ROI\_section\_minus1**[ i ] | u(v) |
| for( j = 0; j <= num\_tiles\_in\_ROI\_section\_minus1[ i ]; j++ ) |  |
| **tile\_index**[ i ][ j ] | u(v) |
| } |  |
| **motion\_vectors\_over\_pic\_boundaries\_flag** | u(1) |
| **max\_bytes\_per\_pic\_denom** | ue(v) |
| **max\_bits\_per\_mincu\_denom** | ue(v) |
| **log2\_max\_mv\_length\_horizontal** | ue(v) |
| **log2\_max\_mv\_length\_vertical** | ue(v) |
| } |  |
| } |  |

# Proposed text change of option 2: using SEI to signal ROI tile sections)

|  |  |
| --- | --- |
| sei\_payload( payloadType, payloadSize ) { | Descriptor |
| if( payloadType = = 0 ) |  |
| buffering\_period( payloadSize ) |  |
| … |  |
| else if( payloadType = = 132 ) |  |
| subpic\_timing( payloadSize ) |  |
| else if( payloadType = = xxx ) |  |
| ROI\_tile\_section ( payloadSize ) |  |
| else |  |
| reserved\_sei\_message( payloadSize ) |  |
| if( !byte\_aligned( ) ) { |  |
| **bit\_equal\_to\_one** /\* equal to 1 \*/ | f(1) |
| while( !byte\_aligned( ) ) |  |
| **bit\_equal\_to\_zero** /\* equal to 0 \*/ | f(1) |
| } |  |
| } |  |

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

# References

1. B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, T. Wiegand, High efficiency video coding (HEVC) text specification draft 8, Document no. JCTVC-J1003, July 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.

# Patent rights declaration(s)

**InterDigital Communications, LLC 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).**