Joint Collaborative Team on Video Coding (JCT-VC)
of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11
15th Meeting: Geneva, CH, 23 October – 1 November 2013

Document: JCTVC-O0045

Title:	RExt: minimum chroma TU size restriction for low-fidelity coding mode
Status:	Input Document to JCT-VC
Purpose:	Proposal

Author(s): David Flynn — *dflynn@blackberry.com*
Nguyen Nguyen — *nnguyen@blackberry.com*
Dake He — *dhe@blackberry.com*

Source: BlackBerry Ltd.

Abstract

In HEVC 4:2:0 coding, the maximum possible depth of the luma and chroma quadtrees are of different by virtue of the chroma components being reduced in resolution by a factor of two and a lack of 2×2 transforms for chroma. This contrasts with the 4:4:4 coding modes where the maximum possible depths are identical. This contribution suggests a method for changing the behaviour of the 4:4:4 coding mode to allow a CU-adaptive restriction on the minimum TU size for a chroma block. Results are presented that show the effect of enabling the behaviour without restriction and of providing an adaptive mode decision mechanism.

This contribution is an enhancement to JCTVC-O0043 that provides a CU adaptive chroma QP adjustment tool.

1 Summary of previous work

In [1] an observation is made that there is a bitrate increase associated with coding an upconverted 4:2:0 video sequence in 4:4:4. In [2] analysis is performed on the observation and a method is presented that attempts to address a case where 4:2:0 content is upconverted and encoded as 4:4:4 using a combination of per CU chroma QP offset adjustment and a minimum TU size restriction.

One of the contributing factors to the observed bitrate increase is derived from the relative increase in chroma block sizes that permits the encoding of what would have been 2×2 TUs in 4:2:0 as 4×4 TUs in 4:4:4.

In HEVC 4:2:0 coding, the minimum TU size for luma and chroma is 4×4. However, since there is a ratio of two in the luma and chroma TU sizes, it is not possible to represent a single co-located chroma TU for a 4×4 block. Rather than introduce a 2×2 TU (and possible 2×2 prediction modes), HEVC refrains from splitting the 4×4 chroma TU of the previous level in the RQT hierarchy. As a consequence, the Intra N×N prediction mode that splits an 8×8 CU into 4×4, only applies to the luma component.

In the current RExt draft, since there is no scaling ratio between luma and chroma dimensions for 4:4:4 coding, an extra level of hierarchy in the RQT tree with is created to represent 4×4 chroma TUs that are co-located with 4×4 luma TUs that incurs an associated signalling overhead. Since there is no restriction in the RQT, there is consequently no need for the Intra N×N prediction mode for 8×8 CUs to apply only to luma.

Another way of viewing this is that the if one were to consider scaling the 4:2:0 system to produce a 4:4:4 codec, the smallest TU size for chroma would be 8×8, with all the 4:2:0 restrictions retained.

The initial RExt investigations [**?**] examined the minimum chroma TU size restriction and the intra N×N restriction (it is not necessary to resolve both).

2 Observations

Table 1 shows that when coding 4:2:0 content unconverted as 4:4:4, there can be a benefit to applying a restriction on the minimum TU size for chroma (and consequentially disabling the N×N Intra prediction mode for chroma). However, for 4:4:4 content, this benefit does not persist, implying that some form of adaptation is required.

Table 1 – BD-rate performance of forcing a minimum TU size restriction when coding 4:4:4 content upconverted from 4:2:0 and 4:4:4 native content.

		BD-Rate		
Configuration	Runtime (%)	Luma	Cb	Cr
i_main-classB-mt		−0.37	−1.02	−1.14
i_main-classC-mt		0.28	−1.13	−1.04
i_main-classD-mt		−0.06	−1.65	−1.79
i_main-classE-mt		−0.77	−3.38	−3.35
i_main-classF-mt		−0.04	−0.87	−1.37
ld_main-classB-mt		−0.28	0.82	1.06
ld_main-classC-mt		−0.34	1.53	1.18
ld_main-classD-mt		−0.52	2.09	2.3
ld_main-classE-mt		−0.3	−0.74	−1.34
ld_main-classF-mt		−1.09	−0.34	−0.72
ra_main-classB-mt		−0.29	0.53	0.7
ra_main-classC-mt		−0.42	1.15	1.06
ra_main-classD-mt		−0.36	2.21	2.23
ra_main-classE-mt		−0.44	−2.01	−1.79
ra_main-classF-mt		−0.22	−0.02	−1.26
ra_rext-yuv444-mt		−1.92	4.23	2.31
ld_rext-yuv444-mt		−2.11	4.08	2.44
i_rext-yuv444-mt		−0.27	2.81	0.16

3 Methods of adaptation

It would be possible to provision control over such a restriction at the PPS or slice level. However, doing so requires an encoder to either perform some form of per-frame pre-analysis (with a consequential buffering delay) or apply some metric determined during the encoding of a previous picture to the encoding of the next. A finer grained control mechanism permits an encoder to amortize the cost of any such decisions over the entire picture.

3.1 Per CU flag

The proposed method is a CU-adaptive scheme that permits an encoder to identify low fidelity areas of the picture, and in each area, changes the implied structure to be better tuned for low fidelity source data.

More specifically, for each CU, an encoder has the option of signalling (or deriving) a flag that, when asserted, has the following effect:

- Applies a restriction to limit the minimum transform unit size for chroma to 8×8. This is analogous to the minimum TU size for chroma of 4×4 in 4:2:0.
- For an 8×8 CU, to signal a single intra chroma mode as opposed to four modes, effectively disabling the N×N split for chroma (as per 4:2:0).

The depth in the coding tree at which the flag may be signalled is tunable using a PPS parameter. When the CTB tree reaches this maximum depth, a single flag is signalled that applies to all CUs below the maximum signalling depth.

3.2 Encoder implementation

A non-normative encoder has been developed that, during the mode decision process, for each CU estimates the transform domain activity of the source data to determine the mode decision. In the case where the current CU is below the maximum depth at which the mode can be signalled, the estimation is performed once for the area covered by the minimum CU size and applies to all smaller CUs in the sub-tree.

The estimator uses the HM transform to analyse the co-located chroma blocks in the source video data to determine the amount of high-frequency activity. The mode decision is based upon a simple threshold that counts the number of high-frequency coefficients with a magnitude above a particular value:

$$est_c = \text{loFiCountThreshold} < \left(\sum |x_{i,j}| > \text{loFiQuant}\right), \text{where } \begin{matrix} x = T(blk_c) \\ i, j \geq \frac{blkSize}{2} \end{matrix}$$

The estimation is performed for both chroma components, and if either exceeds the threshold, the mode is asserted. The thresholds for loFiQuant and loFiCountThreshold are 8 and 8 respectively.

Undoubtedly better estimators or mode decision processes may be developed, including, for example, a full RDO process to test the relative merits of the mode.

3.3 Combined method with chroma QP adjustments

The method presented is suitable as either a stand-alone method or a may be merged with the chroma QP adjustment techniques of [3] or as per the original [2] from which this method is derived.

In doing so it is recommended to provide independent mechanisms to enable these two techniques.

When both chroma QP adjustment and the TU restriction are enabled, the TU size restriction can be inferred from the decision to apply a chroma QP adjustment. The only necessary syntax modification is that the per CU flag that indicates that an adjustment must occur prior to the intra prediction mode signalling in order to limit the number of chroma intra prediction modes signalled for the N×N case and to know the maximum chroma quadtree depth.

4 Results

Table 2 shows results for the cu-adaptive case where a fidelity estimator is used to control the mode selection.

Table 2

		BD-Rate		
Configuration	Runtime (%)	Luma	Cb	Cr
i_main-classB-mt		−0.17	−0.39	−0.41
i_main-classC-mt		−0.02	−0.34	−0.32
i_main-classD-mt		−0.1	−0.39	−0.44
i_main-classE-mt		−0.71	−3.19	−3.35
i_main-classF-mt		0	−0.29	−0.49
ld_main-classB-mt		−0.09	0.34	0.32
ld_main-classC-mt		−0.06	0.19	0.06
ld_main-classD-mt		−0.19	0.54	0.62
ld_main-classE-mt		−0.35	−1.16	−1.23
ld_main-classF-mt		−0.24	−0.46	−0.41
ra_main-classB-mt		−0.08	0.02	0.01
ra_main-classC-mt		−0.07	−0.11	−0.09
ra_main-classD-mt		−0.02	0.69	0.93
ra_main-classE-mt		−0.39	−1.95	−2.37
ra_main-classF-mt		0.06	0.02	−0.63
ra_rext-yuv444-mt		−0.02	0.04	0.13
ld_rext-yuv444-mt		0	−0.01	−0.03
i_rext-yuv444-mt		0	0	0

5 Specification Text

Please see enclosed document for specification text.

BlackBerry Limited 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] M. Budagavi and M. Zhou, "Mixed chroma format coding," JCTVC-L0250, JCT-VC, Jan. 2013.

[2] D. Flynn, N. Nguyen, and D. He, "Rext: Fidelity adaptive coding mode," JCTVC-N0292, JCT-VC, Jul. 2013.

[3] D. Flynn, "Best-effort decoding of 10-bit sequences [missing]," JCTVC-O0043, JCT-VC, Oct. 2013.