Joint Collaborative Team on Video Coding (JCT-VC)
of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document: JCTVC-O0337
15th Meeting: Geneva, CH, 23 October – 1 November 2013

Title: Caution required when evaluating the performance of screen content sequences

Status: Input Document to JCT-VC

Purpose: Information

Author(s): David Flynn | *dflynn@blackberry.com*
Dake He | *dhe@blackberry.com*
Gaëlle Martin-Cocher | *gmartincocher@blackberry.com*

Source: BlackBerry Ltd.

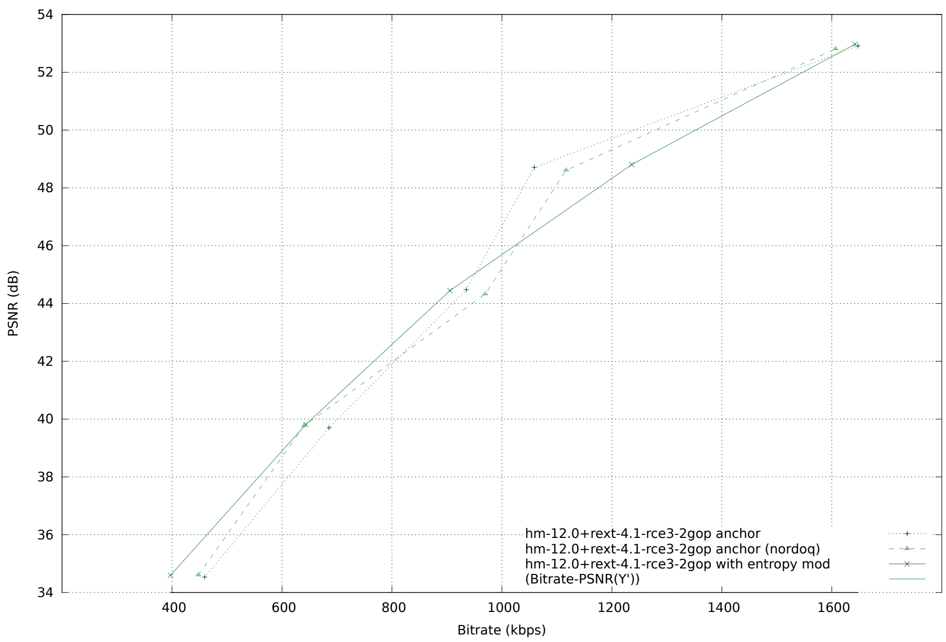
Abstract

Screen content sequences are typified by periods of long global inactivity combined with sudden large scale changes. While inactivity may be efficiently compressed from one frame to the next, sudden large changes can take a considerable number of bits to encode. When sudden changes due to large motion displacements occur, encoder performance can become meta-stable. That is, a minor and otherwise benign coding decision change can lead to large scale coding performance swings. This contribution analyses one such case present in the RCE3 anchor sequences and illustrates the effect observed.

1 Example

In the RCE3 common test conditions [1], interesting effects have been observed by the authors for some sequences. By way of example, the rate-distortion graph for 64 frames of the screen content video conferencing doc sharing sequence is presented for various encoder configurations in Figure 1.

Figure 1 – Rate-distortion graph from encoding sc_video_conferencing_doc_sharing_1280x720_30_8bit_-300_444_r1.yuv at QPs 37, 32, 27, 22 and 17 using a low-delay configuration



Examining the anchor plots (dotted lines), the graph looks a little peculiar, containing several kinks. At a first glance there is a remarkable performance drop in the anchors between QPs 22 and 27, but when one examines the two anchors in further detail even more bizarre behaviour is observed. The two anchors differ only by their use of RDOQ; the RCE3 anchor has RDOQ enabled, while the alternative anchor has it disabled[1]. Referring again to the graphs, the version without RDOQ is indicated to be substantially

[1] `--RDOQ=0 --RDOQTS=0`

better than the anchor with at QP 32, and to a lesser extent at QPs 37 and 17. Typically one would observe the anchor without RDOQ to perform with approximately a constant detriment in performance to that with it enabled.

Furthermore, when a minor change is made to the entropy coder one observes yet again drastic changes in the rate-distortion curve.

2 An examination at QP 32

To provide some insight into the observation made from the rate-distortion graph, the encoding performance and encoder decisions are examined on a per-frame basis from a sub-sequence of the encoded bitstreams at QP 32.

The sub-sequence comprising of frames 36–42 contains a sudden large-scale change. Figure 3 plots the per-frame PSNR and compressed frame sizes for this sub-sequence, showing that frames 39 and 40 may be examples of the differing behaviour.

Pertinent frames from the sub-sequence are shown in Figure 2. Frames previous to, and including, 38 are essentially identical, with no significant variation in picture content. At frame 39 the main application window is scrolled rapidly from left to right. The motion is so rapid that tearing is observed partway through the window – the nature of the tearing is an effect of the application rather than the video capture method. Three regions can be identified; those with zero motion, those with, relative to frame 38, a horizontal displacement of 244 pixels, and those with a horizontal displacement of 541 pixels. Frame 40 concludes the scrolling action (at 541 horizontal pixels relative to frame 38) and subsequent frames continue to show little variation.

A mode analysis of the encoded bitstreams for frames 38, 39 and 40 is presented in Figure 4.

Frame 38 is, as expected, not particularly interesting. It consists entirely of skipped CTUs, which incidentally reveals the cost of a skipped 720p frame is around 112 bits. All three bitstreams behave identically.

Frame 39 is largely coded as intra, which is not too surprising given that between a third and a half of the frame is revealed for the first time. In this region, there are minor mode changes between the three bitstreams as would be expected from any minor perturbation to the encoder state. The right hand side of the window is more interesting, however. The number of skipped CTUs is significantly different in the three sequences. Comparing Figures 4d and 4e, the latter has more skipped CTUs: approximately eight in the case of the first CTU row.

Frame 40 shows the same behaviour on the right hand of the window again with significant variation between the three sequences. In conducting various experiments, the authors have seen even larger variations than presented here.

2.1 Probable cause

The mode figures reveal a reluctance to use the skip mode. Or perhaps more accurately, it should be noted how the use of the skip mode propagates through the picture. Once it has been selected, subsequent blocks seem to have a higher probability of also being skipped, resulting in large areas that correspond to the areas of homogeneous motion.

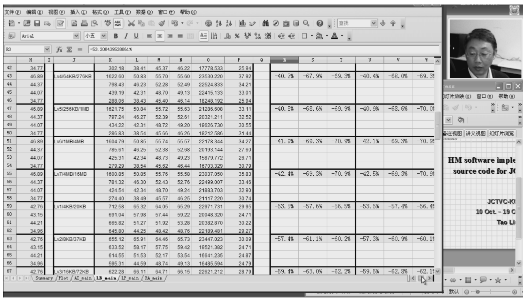
Such behaviour could be an effect of the greedy mode search employed by the HM, where a given block is solved for all modes with no consideration for subsequent blocks. Given that there is a requirement for skipped blocks to use a motion vector prediction, it is very possible that the cost of coding the delta motion vector and inter mode for a previous block is similar to that of the intra mode. As such, the encoder can become caught in a local minimum, which is only broken by the noise inherent in the mode decision process.

3 Concluding remarks

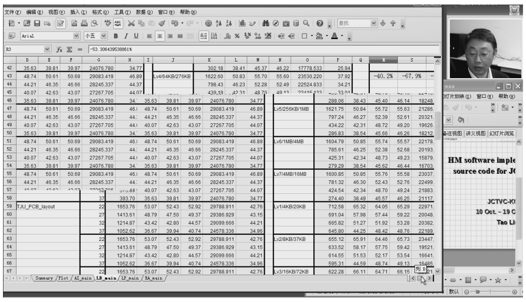
Sequences showing such effects can be highly unreliable indicators of intrinsic coding tool performance, exhibiting large arbitrary positive or negative swings in BD-rate performance. Care should be taken when interpreting their results as a gain may be indicated, whereas in the optimal case a loss should be reported.

Figure 2 – Still pictures from the video conferencing doc sharing sequence.

(a) Frame 38



(b) Frame 39



(c) Frame 40

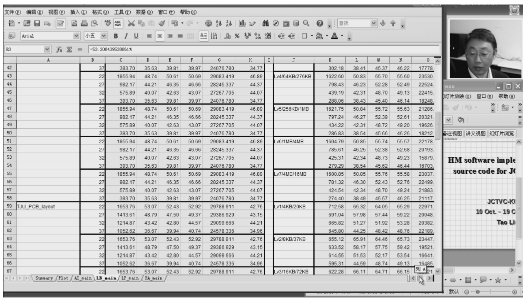
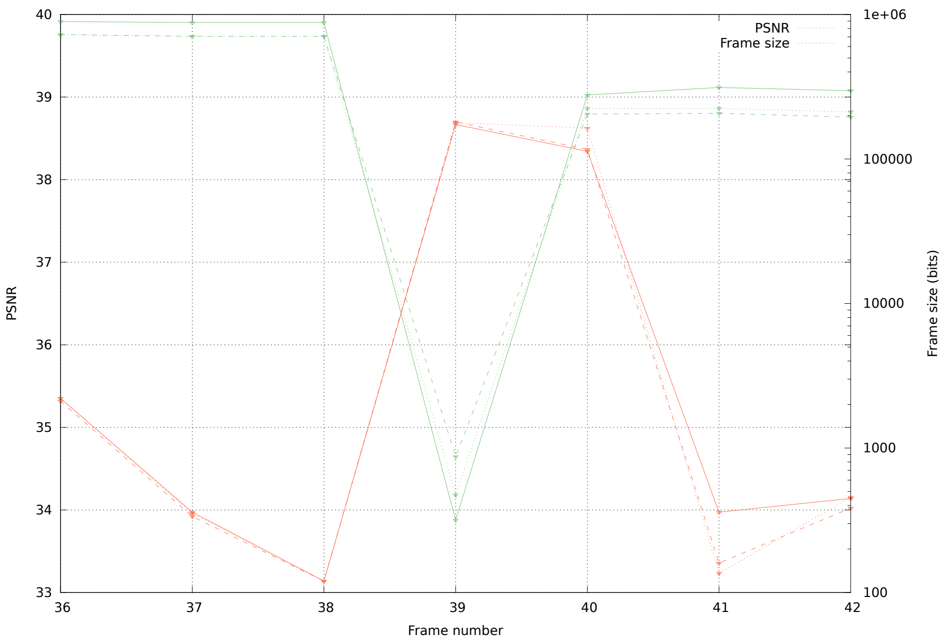
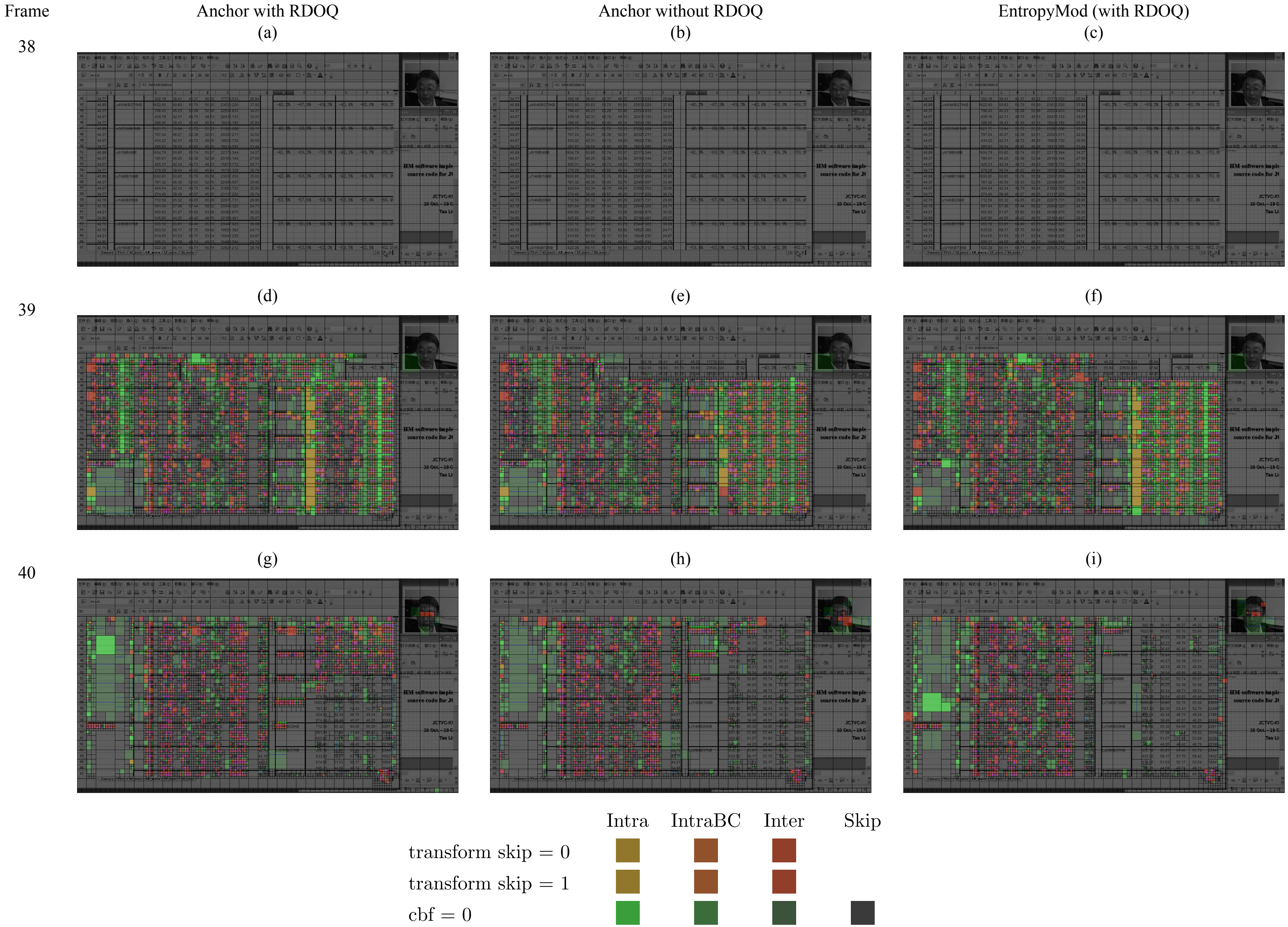


Figure 3 – Per-frame PSNR and compressed bitrate plots for video conferencing doc sharing sequence at QP 32 using a low-delay configuration.



Date saved: 2013-10-23

Figure 4 – Frames 38–40 of the video conferencing doc sharing sequence at QP 32 using a low-delay configuration with mode usage information overlayed.



Date saved: 2013-10-23

References

[1] A. Saxena, D. Kwon, M. Naccari, and C. Pang, "Hevc range extensions core experiment 3 (rce3): Intra prediction techniques," JCTVC-N1123, JCT-VC, Jul. 2013.