Rate-Distortion Performance Of Contemporary Video Codecs:

Comparison Of Google/WebM VP8, AVC/H.264 and HEVC TMuC

Etito Ohwovoriole and Yiannis Andreopoulos

University College London

Abstract: This paper gives an experimental primer on the current affairs in state-of-the-art video compression, focusing purely on rate-distortion performance under quantization or bitrate constraints and disregarding system complexity, delay and other domain-specific factors. Our comparison includes: the Advanced Video Coding (AVC/H.264) standard as instantiated by the VideoLAN x264 project, the VP8 codec as provided by the Google/WebM project and the High Efficiency Video Coding (HEVC) Test Model under Consideration (TMuC) that is currently under intensive development for joint standardization by ISO/IEC and ITU-T. We use six standard-definition progressive-scan test video sequences for our comparison, which correspond to video content that can be streamed to/from a variety of systems, including mobile devices. For the same video distortion (in terms of peak signal-to-noise ratio and structural similarity metric), our comparison reveals that, on average, HEVC TMuC v0.5 provides 46% bitrate reduction in comparison to AVC/H.264 (x264 v0.85.xx, optimizing for PSNR), which in turn provides 21% bitrate reduction in comparison to Google/WebM VP8.

1. Introduction

Video compression benchmark tests [1] are important for assessing the operational rate-distortion (RD) performance of video encoding and decoding algorithms (video codecs). Given that video streams tend to dominate other network flows in terms of bandwidth requirements, such tests allow for informed decisions on technology deployment and bandwidth provisioning for media-oriented networking systems. In this paper, we present a comparison of video coding schemes that have (or are expected to have) prevailing importance in video transmission systems within this decade.

Several video codec comparison summaries have appeared in the public domain recently[1]. The main focus of such tests is standard-definition or high-definition material, as this is the primary video source for mainstream services (e.g. YouTube, BBC iPlayer) today. Such tests tend to include the latest open-source video codecs based on well-established ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) series of standards. Of particular relevance to contemporary video services are:

- the MPEG-VCEG AVC/H.264 standard [2] as provided by VideoLAN's x264 project;
- the Google-WebM VP8 coder [3], a variant of the standardized MPEG/VCEG schemes that includes some, arguably more-advanced, prediction features aimed (primarily) at low bitrates;
- the upcoming High Efficiency Video Coding (HEVC) standard, currently under intensive joint development by ISO/IEC MPEG and ITU-T VCEG [4].

Our motivation for this work stems from the following:

- Even though the existing RD comparisons are informative, they are not without problems when quantifying comparative performance between different codecs. Such problems relate to: the use of already-compressed video material that contains compression artifacts, the use of non-standard metrics for quality comparison, the reporting of results under different optimization settings for each coder and under the use of external tools to the video coding scheme (such as: tools for scene-change detection, pre- or post-processing), and the use of ad-hoc summative assessment metrics.
- On the other hand, video coding comparisons done during a standardization process, albeit following a methodology established by experts [1][5], tend to avoid practical instantiations of video codecs such as VideoLAN x264 and WebM VP8 because of their evolving state and simply

---

This work was supported by EPSRC EP/F020015/1.

[1] e.g.: http://www.quavlive.com/video_codec_comparison, http://x264dev.multimedia.cx/?p=377, http://www.compression.ru/video/codec_comparison/index_en.html

utilize the AVC/H.264 reference software [1][2], something that is not informative on the practical performance benefit against contemporary state-of-the-art developments in the area.

- Finally, to the best of our knowledge, there is currently no joint comparison between all three schemes identified above.

The aim of this work is to perform such a comparison. We focus on the high profile of each of the three codecs[2] named above, with all possible optimizations for maximum quality under quantization or bitrate constraints. This is targeting the "highest quality" profile, which is commonly chosen by storage or bitrate-conscious content providers. We utilize six uncompressed standard definition video sequences available in YUV 4:2:0 (raw) format[3]. The summary of our results ranks the measured performance in terms of percentile bitrate reduction under the same peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index, based on the Bjontegaard measurement method [5].

2. Outline of Codecs under Comparison

The ISO-IEC/MPEG ITU-T/VCEG AVC/H.264 standard is a very successful video coding scheme that, for the same decoded video distortion, reduces bitrate requirements by approximately 50% [1][2] in comparison to MPEG-4 or H.263 standards. This is achieved by an arsenal of tools [2], including multi-frame variable block-size motion-compensated prediction, advanced context-based entropy coding, advanced temporal prediction structure including hierarchical bi-directional prediction modes and an adaptive in-loop deblocking mechanism for improved visual quality at low bitrates. The open-source implementation of VideoLAN x264 is currently at an advanced development stage and can be said to obtain comparable RD performance[2] to the best commercially-deployed AVC/H.264 solutions.

The recent open-source development from Google/WebM, i.e. VP8, includes very similar mechanisms, but with somewhat simplified deblocking mechanism and limited possibilities for temporal prediction structures [3] in comparison to AVC/H.264. As an alternative, VP8 employs an adaptive mixing strategy for creating artificial reference frames [3] (so-called "golden frames"), which is aimed at providing improved coding efficiency for low-bitrate coding. The current codec implementation still has significant room for improvement by extending the current specification to include more advanced interpolation, deblocking, entropy coding and temporal prediction features.

Seeing the explosion of high-definition content and user-generated content that necessitates advanced coding methods for bitrate reduction, the Joint Video Team that developed AVC/H.264 evolved into a new team, the Joint Collaborative Team on Video Coding (JCT-VC) of MPEG and VCEG. JCT-VC was created to[4]: "develop a new generation video coding standard that will further reduce by 50% the data rate needed for high quality video coding, as compared to the current state-of-the-art AVC standard (ITU-T H.264 | ISO/IEC 14496-10)". The work is currently under the Test Model under Consideration (TMuC) stage, which, as indicated by its name, is currently undergoing exhaustive testing and tool refinement [4]. TMuC is expected to converge to the Test Model (TM) v1.0, whose further refinements will converge towards the final ISO/IEC ITU-T international standard.

3. Experimental Setup

a. Test Material and Rate-Distortion Measurement

Our test material consists of the YUV 4:2:0 sequences[3] "Crew", "Harbour", "Raven", "Soccer", "Sailormen" and "Shuttlestart". All sequences contain 600 video frames, each frame consisting of 704x576 pixels for the Y channel and 352x288 pixels for the U, V channels in progressive (raster) scan. The coding and playback frame-rate is kept constant at 30 frames-per-second. PSNR measurements for each test were averaged over all frames (creating $\text{Mean\_PSNR}_{\{Y,U,V\}}$ per sequence, codec and encoding point) for the luminance and chrominance channels, in order to quantify log-scale mean-square error against the original video. We use a single metric per point:

$$\text{Mean\_PSNR} = (4 \times \text{Mean\_PSNR}_Y + \text{Mean\_PSNR}_U + \text{Mean\_PSNR}_V)/6$$

---

[2] x264 v0.85.xx: http://www.videolan.org/developers/x264.html; WebM VP8: http://www.webmproject.org/; HEVC TMuC 0.5 build 83: svn://hevc.kw.bbc.co.uk/svn/jctvc-tmuc/ accessed July 28, 2010.

[3] http://media.xiph.org/video/derf/

[4] http://www.itu.int/ITU-T/studygroups/com16/jct-vc/

in order to give higher emphasis on the luminance channel but also include chrominance-channel PSNR in our results. This was because we found that VP8 tends to underscore in the chroma channels for each sequence in comparison to x264 and TMuC.

Besides PSNR, we utilized the SSIM index of Bovik *et al* [6] as a quantitative measure of visual quality against the original (uncompressed) content. SSIM provides unit-less values between 0.0 and 1.0; results above 0.95 correspond to decoded video that tends to be visually-indistinguishable from the original, while scores below 0.8 correspond to easily-distinguishable visual artifacts. The produced SSIM indexes were scaled by 100 and averaged for the Y channel per sequence, codec and encoding point by using the provided software [6], thereby creating Mean_SSIM (%) for each point.

Finally, the average bitrate (reported in kbps) per sequence, codec and encoding point was measured based on the actual bitstream size for each (entire) video sequence and the utilized frame-rate.

b. Codec Settings

Both x264 and TMuC are set to quantizer-based control, with the selection of base quantization value (QP) as: QP ∈ {22, 26, 30, 34, 36, 38, 40, 42, 44}. Concerning VP8, we did not find a recommended approach to allow for quantizer-based control and, as a result, we encoded each video sequence with two-pass variable-bitrate control, set at bitrates matching the x264 average bitrates.

For all coding schemes, all advanced temporal prediction features and entropy coding tools were enabled as recommended by their developers: high profile and --placebo for x264, --best for VP8, recommended parameters [4] for maximum performance in TMuC v0.5. Fixed intra refresh rate every 64 frames was imposed in all schemes in order to allow for random access. We deliberately did not use scene-change detection for adaptive selection of intra frames, as: *(i)* this can be implemented outside the core video coding framework and *(ii)* its success varies according to how suitable is the detection process to the test video sequences' characteristics. Both x264 and TMuC utilize hierarchical (pyramidal) B-frame temporal decompositions under the recommended settings for maximum performance. VP8 does not support such a feature; it compensates via the use of artificial "golden" reference frames created by reference-frame mixtures instead. Finally, specifically for the case of x264, we create a second set of results with the same settings, plus the flag --tune PSNR. This is configuring internal operation to optimize for PSNR, which is expected to allow for adaptive QP selection. PSNR-oriented encoder optimization is performed by default within VP8 and TMuC.

All available settings, compressed bitstreams and references for utilized binaries for encoding and decoding are made available online at the related webpage[5]. Despite our best effort to create equivalence in the utilized settings, we would like to stress that the reported experiments express the snapshot of each codec's development at the time of this writing.

4. Results

Indicative RD results are given in Figure 1. For each codec and each sequence we performed cubic polynomial fitting of the log (base-10) bitrate vs. PSNR results, or log (base-10) bitrate vs. SSIM percentile results. This corresponds to the Bjontegaard Distortion-rate (BD-rate) measurement [5], via which we calculated the average bitrate reduction of each codec under equal PSNR and under equal SSIM per sequence. We report the results in Table 1. VP8 appears to be performing a reconfiguration at the low bitrate regime, which makes it outperform x264 for a limited range of bitrates. An example of this can be seen at the low-PSNR and low-SSIM regime of Figure 1. This effect leads to significantly-improved quality for the low-bitrate regime of our tests. We remark here that this irregularity of VP8 (along with the usage of two-pass rate control) makes the Bjontegaard method somewhat questionable for its assessment. The overall PSNR-log(bitrate) or SSIM-log(bitrate) curve of VP8 could be piecewise-approximated per sequence, in order to cover the particular region of low-bitrate measurements separately from the rest.

PSNR and SSIM assessment seem to be in agreement with each other for the majority of the results. While x264 is a respectable 20% superior to VP8 in BD-rate sense, HEVC has potential to allow for significant bitrate savings (potentially approaching 50%) for the same video quality as AVC/H.264.

[5] http://www.ee.ucl.ac.uk/~iandreop/CCOMP.html

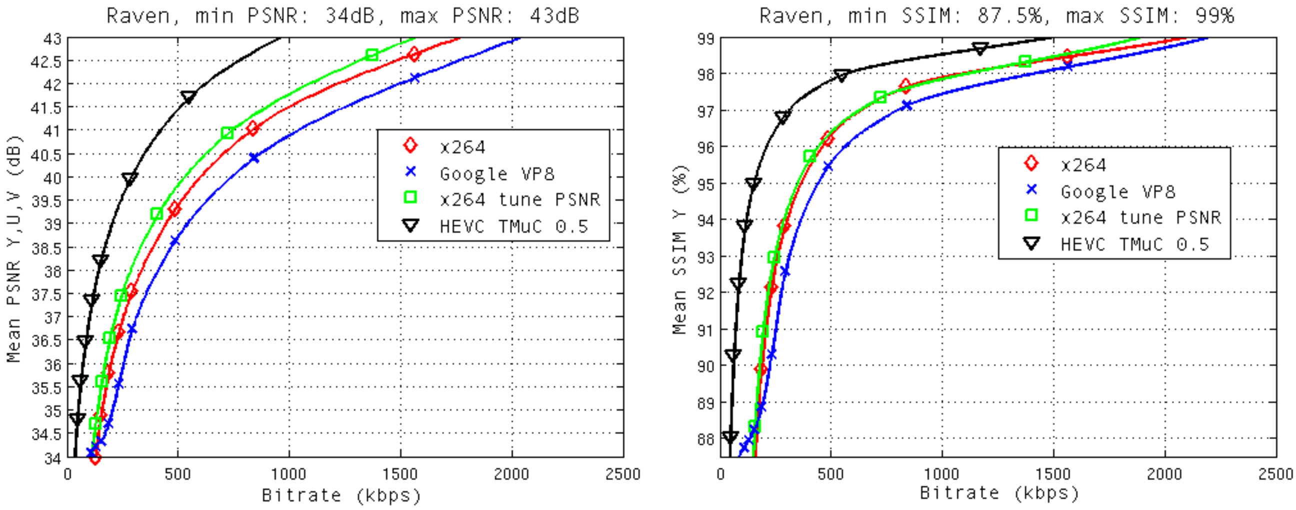

Figure 1. Rate-distortion results of sequence "Raven". Left: PSNR results, right: SSIM results. The points represent measurements while the curves have been fitted to the experimental data by cubic interpolation within the specified range of PSNR or SSIM.

Sequence	Bitrate reduction (%) under equal PSNR (across the range) based on the RD fit		Bitrate reduction (%) under equal SSIM (across the range) based on the RD fit	
	x264 vs VP8	TMuC vs x264	x264 vs VP8	TMuC vs x264
Crew	9.07	40.20	9.36	46.21
Harbour	31.95	32.76	25.08	51.82
Raven	27.58	51.89	20.03	56.73
Soccer	22.68	37.94	13.63	40.22
Sailormen	31.19	46.47	24.95	44.41
Shuttlestart	24.65	52.47	14.46	51.79
Average:	**24.52**	**43.62**	**17.92**	**48.53**

Table 1. Average percentile bitrate reduction under the same PSNR and SSIM [5]; "x264" corresponds to the x264 experiments with the --tune PSNR setting and "TMuC" to HEVC TMuC v0.5.

5. Conclusion

Our comparison of three contemporary video coding schemes in terms of rate-distortion is summarised in Table 1. Given that HEVC TMuC already has significantly-higher RD performance while still being under heavy development and testing, its future instantiations (either within the future standard or within x264 or VP8 as a battery of optional "high-efficiency" tools) may begin to replace AVC/H.264 and VP8 in R&D efforts within the next 2 years and within products and services within 3~5 years.

References

[1] P. Lambert, *et al*, "Rate-distortion performance of H.264/AVC compared to state-of-the-art video codecs," *IEEE Trans. on Circ. Syst. for Video Technol.*, vol. 16, no. 1, pp. 134-140, Jan 2006.
[2] T. Wiegand, *et al*, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circ. and Syst. For Video Technol.*, vol. 13, no. 7, pp. 560-576, July 2003.
[3] *VP8 data format and decoding guide*, Google Inc., WebM project, July 2010.
[4] F. Bossen, *et al*, "AHG report: Software development and TMuC software technical evaluation," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 WP3* (JCT-VC), JCTVC-B003, July 2010.
[5] G. Bjontegaard, "Improvements of the BD-PSNR model," *ITU-T SG16*, VCEG-AI11, Jul. 2008.
[6] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Process.*, vol. 13, no. 4, pp. 600-612, April 2004.