

**ISO/IEC JTC 1/SC 29/WG 11**

**Coding of moving pictures and audio**

**Convenorship: UNI (Italy)**

**ISO/IEC JTC 1/SC 29/WG 11 N18N19160**

**Document type: Approved WG 11 document**

**Title: Test results of LTM 3.0**

**Status: Approved**

**Date of document: 2019-02-14**

**Source: Video**

**Expected action: N/A**

**No. of pages:**

**Email of convenor: leonardo@chiariglione.org**

**Committee URL: mpeg.chiariglione.org**

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/WG 11 N19160**

**Brussels, BE – January 2020**

|  |  |
| --- | --- |
| **Source:** | **Video** |
| **Title:** | **Test Results of LTM 3.0** |
| **Authors :** | **Vittorio Baroncini (Test Chair), Stefano Battista (UNIVPM), Lorenzo Ciccarelli (V-Nova), Simone Ferrara (V-Nova), Thomas Guionnet (ATEME), Guido Meardi (V-Nova)** |

**Test Results of LTM 3.0**

**Abstract**

During the 129th MPEG meeting the LCEVC BOG released the LCEVC Test Model v. 3.0.0 reference software. Subsequently, a debugged version 3.0.1 was released.

This document describes the results that have been obtained by testing LTM 3.0.1 (“LTM 3.0.1”) under the LCEVC Common Test Conditions as described in [1] according to the guidelines and configuration files described in Annex 1 “Instructions for how to reproduce the CTC testing results” of [2].

Preliminary MOS results[[1]](#footnote-2) were obtained using the DSIS methodology, as recommended by the ITU-R Recommendation BT.2095, under supervision of the Test Chair. Larger-scale formal MOS testing with confidence intervals will be executed on the upcoming LTM 4.0.

This document provides current CTC results prepared by V-Nova (Section 2) and cross-checks of these results by ATEME (Section 3).

1. **Characteristics of LCEVC coding scheme and implications for measuring efficiency benefits**

Since the low complexity characteristics of LCEVC are relatively evident (and were further explored during Core Experiment 2 [3]), much time has been invested on determining the most appropriate way to assess the coding efficiency benefits that LCEVC brings to an enhanced codec.

In particular, LCEVC requires to compare a traditional coding structure with a “layered” coding structure leveraging non-overlapping transform sizes as well as scaling operations, which is known to decrease the reliability of objective metrics [4 et al.]. Comparing different video codecs with simple objective metrics implicitly assumes that the tested codecs treat pixels in a similar way: this is not the case for LCEVC, and not all pixels are created equal.

In fact, LCEVC does not necessarily perform badly with metrics. Simply, objective metric BD-rates often diverge materially across different metrics (e.g., PSNR vs. VMAF), and subjective assessments often diverge from conclusions that can be drawn by just looking at metrics. For all tests conducted so far by the proponents and/or the BOG, formal subjective assessments double-checked with expert viewing have showed materially higher BD-rate benefits than it was suggested by objective metrics.

For example, results of the LCEVC CfP as well as test results provided by proponents in the past (e.g., [2] [5] [6] et al.) suggest that VMAF scores are closer to subjective scores than PSNR, likely due to VMAF having been designed to compare the quality of encodes at different resolutions (and thus it is more «accustomed» to scaling operations). However, also VMAF scores seem to be penalizing for LCEVC with respect to actual MOS scores.

1. **Current CTC performance results for LTM 3.0.1 Reference Software**

Table 3, 4 and 5 below show the results[[2]](#footnote-3) obtained by the LCEVC Reference Software LTM v.3.0.1 when running the LCEVC CTC [1] using Configuration A as per [2] Annex 1.

Table 1 **– UHD enhancing JM**





Table 2 **– UHD enhancing HM**





Table 3 **– HD enhancing JM**





Table 2-5 above also show the RD curves for each comparison. It should be noted that – in line with the requirements of the LCEVC CfP [7] and of Core Experiment 1 [9] – the QPs for CTC anchors were selected so as to span across a wide range of bit rates, from very low quality to very high quality. CTC sequences also include two synthetic and extremely sharp eSports clips.

1. **Cross-checks and verification of test results**

This section presents results and comments provided by ATEME. Tests have been performed using LTM 3.0.1 (Tag LTM\_3\_0\_1 from January 27th). Tests have been conducted following the CTC, as defined in MPEG document N18988. Both anchors and LTM encodings have been performed and checked. Anchors encodings have been found to be almost identical, with minor differences identified. LTM encodings have been found to be near identical, except for one sequence.

Results summary are reported in Table 4, Table 5 and Table 6.

Table 4 – **UHD enhancing JM**

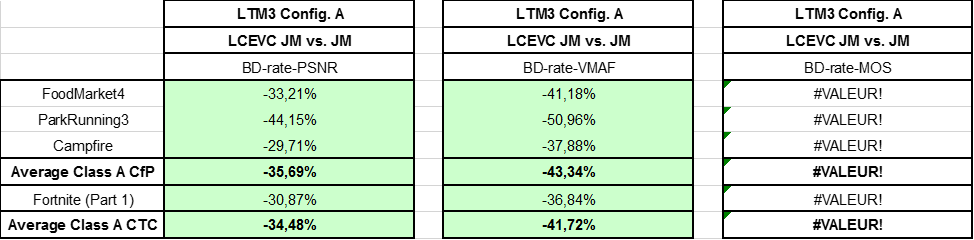
****

Table 5 – **UHD enhancing HM**

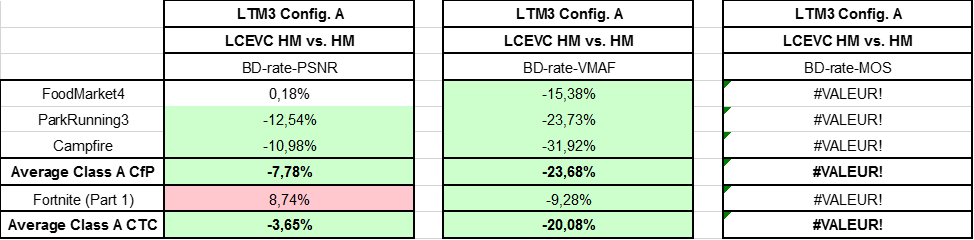
****

Table 6 - **HD enhancing JM**



* 1. **Comments on results**
* UHD enhancing JM (Table 4): proponents and cross-checker results are strictly identical.
* UHD enhancing HM (Table 5): There is a difference in the anchors for the Campfire sequence. It is explained by a mistake in the parameters made in the results reported in Table 2. A QP is shifted by one compared to CTC, thus generating a mismatch. By using the same QP for the results provided in Table 2, the cross-checker obtained identical results to those in Table 2. There is a bitrate difference for the Parkrunning3 sequence between proponent and cross-checker LTM encodings. This is unexplained as of now. However, as metrics scores are identical, and other sequences are fully identical, it is speculated that it is related to a minor mistake. Further investigation will be conducted.
* HD enhancing JM (Table 6): The marginal difference is coming from the anchors. Although this small difference is not explained yet, it is a minor issue, that will be investigated, nonetheless.
* Due to a heterogeneity issue in our computing grid, measured processing times generated were not reliable, and thus are not reported. Software has been inspected, and although no numerical results are provided, the cross-checker would like to mention that no computationally intensive functions were noticed. The encoding/decoding process is fairly simple, especially compared to state-of-the-art codecs.
  1. **Comments on CTC**

As it happened, reproducing anchors and LTM encodings was not straightforward. Numerous issues had to be solved. In this respect, the cross-checker would like to thank V-Nova people for their availability and helpfulness.

It is recommended to revisit the CTC. To this end, the following comments are provided as a basis for discussion.

* Text clarification
  + Having the quantization settings spread between the CTC document and the excel sheet is unpractical. Everything could be in the same place.
  + The configuration files usage must be clarified. Confusions and mistakes happen with the current text, especially for the JM.
  + Examples of command lines could be provided in the CTC document.
  + For easier reading, order of sequences in the tables could be made consistent.
* Quantization parameters
  + As already discussed, quantization parameters are set on a per sequence basis. It allows providing results in use-case relevant bit rate ranges.
  + However, for anchors and base layers, some settings are very similar. Would not it be enough for the specific context of the CTC to have a single set of QP for anchors and base layers?
  + The LCEVC quantization steps (LOQ) are not represented “as usual”. Could it be possible to provide some insight on the meaning of LOQ and its impact on rate/quality?
  + As far as I understand, LOQ maximum value of 32767 means no residual. Does that mean that LOQ of 32767 is equivalent to not having enhancement?
  + For the second enhancement layer, LOQ is always equal to 32767. Does that mean that the second enhancement layer is not tested within the CTC framework?
  + If some parts of the LTM configuration are using only base layer, should not we consider the use of quarter resolution as anchor? Or a convex hull of full and quarter resolutions?
* JM configuration
  + For the specific usage of JM, three configuration files are used: one default, one for HD and one for UHD. This could be simplified.
  + First, by using proper option, one can get rid of the default cfg file.
  + Second, the UHD config file raises issues:
    - The results are worse than with the random access cfg used for HD, which questions the relevance of the test.
    - We are supposed to use the same configuration for anchors and base layers. However, for a UHD anchor, the base layer is HD
  + Finally, I would recommend using a single configuration file for JM, the well-known random access configuration.
* Base versus anchors encoding
  + As already mentioned, base layer is supposed to use the same configuration as corresponding anchor. In the case of the JM, it causes a consistency issue with the configuration file, as the UHD configuration is used for the HD base layer.
  + QP are different between base and anchors. Could it be possible to have some explanations? Is there a rule allowing to fix relative QP between base and anchor?
* LTM software and base layer
  + The LTM provides two ways of handling the base layer:
    - Either one provides the already processed base as a YUV file
    - Or one let the LTM conduct all the process (resize, base encoding, LCEVC encoding, bitstream construction)
  + The latter case is the most practical. However, it is implemented with some hard-coded instructions, namely codecs, configuration files and command lines.
  + It would be useful to revisit the LTM interface in order to make the base layer encoding more visible. For example, passing the configuration file and encoder/decoder path as parameter would avoid doubts on the exact nature of the experiments.

1. **Conclusions**

LTM3 v.3.0.1 shows improved VQ performances against the respective leveraged codecs.

Further improvements are expected with the next versions of the Test Model (LTM v.4.0), which are expected to include additional VQ-enhancing tools not yet leveraged by LTM v.3.0.1.

A revisit of the CTC is recommended in order to ease the testing process and the results analysis.

1. **References**

[1] “Common Test Condition of Low Complexity Enhancement Video Coding”, ISO/IEC JTC1/SC29 WG11 N18988, Brussels, BE, January 2020

[2] “LTM v.3.0.0 – CTC Results”, ISO/IEC JTC1/SC29 WG11 m52268, Brussels, BE, January 2020

[3] “Report on Core Experiment 2”, ISO/IEC JTC1/SC29 WG11 Doc. m49254, Gothenburg, SE, July 2019

[4] https://en.wikipedia.org/wiki/Peak\_signal-to-noise\_ratio.

[5] “Test Chair Report – Results of the Formal Subjective Assessment Test for the CfP for Low Complexity Video Coding Enhancements” ISO/IEC JTC1/SC29/WG11 m47961, Geneva, CH, Mar. 2019.

[6] “[LCEVC] – Performance Characterization over a range of enhanced codecs” ISO/IEC JTC1/SC29/WG11 m51438, Geneva, CH, Oct. 2019.

[7] “Call for Proposals on Low Complexity Video Coding Enhancements”, ISO/IEC JTC1/SC29 WG11 Doc. N17944, Macao SAR, CN, October 2018.

[8] “Requirements for Low Complexity Video Coding Enhancements” ISO/IEC JTC1/SC29/WG11 N18098, Macao, CN, Oct. 2018.

[9] “Report on Core Experiment 1”, ISO/IEC JTC1/SC29 WG11 Doc. m49253, Gothenburg, SE, July 2019

1. Executed on a small set of viewers according to formal DSIS methodology, with expert viewing double-check of results. [↑](#footnote-ref-2)
2. MOS scores are preliminary, and currently being double-checked by the Test Chair [↑](#footnote-ref-3)