**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/AG 5**

**MPEG VISUAL QUALITY ASSESSMENT**

**ISO/IEC JTC 1/SC 29/AG 5 N39**

**Online – October 2021**

|  |  |
| --- | --- |
| **Title** | **Guidelines for Verification Testing of Visual Media Specifications** |
| **Source** | **AG 5 MPEG Visual Quality Assessment** |
| **Status** | **Approved** |
| **Serial Number** | **20975** |

# Introduction

This document defines guidelines for MPEG verification tests which shall serve as a reference for ongoing and future verification test activities. It describes verification testing of standards which define visual media output. In order to assert the suitability of the standard for visual assessment by users, formal subjective evaluation of the visual media signal reconstructed or synthesized from the compressed bitstream is mandatory.

MPEG verification tests has been since the very beginning a fundamental step in the path to the release of a new standard. Nevertheless, the new organization of SC29 in Working Groups and Advisory Groups induces the need for a procedure defining the necessary communication and the exchange of test results between the relevant WG and AG5. The document is supposed to reflect the common understanding of previous verification test activities in MPEG, and the transition of the best practice into the new organizational structure of MPEG.

The description currently relies on the assumption of video sequences to be compressed and reconstructed. For visual media which is rendered (e.g. in case of 3D-video, or 3DoF, 3DoF+, and 6DoF), additional steps may be required and should be added in a revision of this document.

# Purpose and Goals of Verification Tests

MPEG verification tests are conducted on the final specification of a (visual media) standard to assess the fulfillment of the goals of the standardization activity. This includes the demonstration of the achieved compression capability of the standard under the conditions of relevant exemplary application scenarios as well as the successful implementation of other requirements defined at the start of the standardization activity.

To this end, the verification test typically compares the performance of the new standard to the performance of the relevant predecessor(s). The comparison is typically based on the reference software which has been used for developing the standard but other implementations may be considered. The verification tests typically target one or more profiles of the standard may be planned and conducted in a step-by-step fashion to support the assessment of one or more relevant application scenarios for each of the profiles.

# Procedural steps in a Verification Tests

* The WG develops the work item for which a verification test should be conducted and brings it to a final state (FDIS).
* The WG consults with AG5 to produce a verification test plan for the work item. The preparation of the verification test plan is suggested to start before reaching the final state.
* AG5 appoints a test coordinator, to be mentioned in the verification test plan, who is responsible for coordination of the verification tests as well as the selection and instruction of qualified test laboratories to perform the formal subjective assessments.
* The verification test plan is approved by both, AG5 and the WG.
* The laboratories report the results to AG5. Upon approval AG5, communicates the results to the WG;
* The WG produces the verification test report for the work item. The report is submitted for public release.

# Selection of Test Material

For each application scenario which is supposed to be tested, a set of representative test sequences shall be collected. It is strongly recommended to exclude sequences which have already been used in the development phase of the standard. The number of test sequences should be balanced with the overall effort. A set of 4 to 5 test sequences for an application scenario may be considered adequate.

It is recommended to collect hashes (e.g. md5sums) of the uncompressed source sequences to enable unambiguous identification. This is especially helpful if multiple variants of a test sequence are available.

# Preparation of Bitstreams and Rate Points

The verification test typically compares the performance of the new standard to the performance of the relevant predecessor (the anchor). Comparable configurations must be used for both, the anchor and the new scheme. The configurations shall be aligned as closely as possible. All relevant tools available in the anchor shall be used.

It is recommended to define 4-5 rate points, covering the MOS range. Two strategies are possible:

* Matching rates between anchor and new scheme. This allows for direct assessment of the quality improvement of the new scheme relative to the anchor;
* Matching quality of rate points. This supports a more reliable computation of rate savings over the covered rate range using the Bjøntegaard Delta Rate method [3]. Depending on the performance relation between the two schemes, it may be possible to achieve both, rate and quality matching at least for some rate points.

Bitstreams should be generated and crosschecked by independent parties to assert the correctness of the encoding process and the compliance to the defined encoder configurations of both, anchor and new scheme.

The typical procedure for determining rate points includes the definition of candidates based on objective metrics with a suggested preselection w.r.t. suitability for subjective evaluation, followed by dry-run experiments which may include experts viewing sessions or viewing sessions with naïve subjects. Formal subjective assessments with naïve subjects are recommended.

# Conduction of Visual Tests

The formal subjective evaluation is coordinated by a test manager who is appointed by AG5 in agreement with the WG. The test manager is responsible for all required logistic, technical, and design activities in the context of the formal subjective evaluation for the verification test.

The coordination activity of the test manager comprises:

* The selection, the direction, the coordination and the instructions to the test laboratory(ies);
* If and when required, any action necessary to assert qualification of new laboratories as capable to perform state-of-the-art formal subjective evaluations meeting the requirements of the verification test;
* The selection of the evaluation conditions and the assertion of their suitability according to the Verification Test Plan;
* The design, the supervision of the conduction (or conduction) of the formal subjective test.
* The collection and the statistical analysis of the data resulting from the test, to be submitted in a report to AG5;
* Co-editing of the verification test report.

# Reporting

The report of the verification test shall include a description of:

* The verification test activity, including scope, results, and test procedures;
* The test conditions;
* The test procedure;
* The test logistic (e.g. equipment, subjects involved, voting scheme and other details);
* Details of the test results including MOS and CI, plots of the MOS results grouped by test sequences;
* Summary of the BD rates values that allow to provide proof of the bit rate savings.

It is recommended to highlight any specific features of the new scheme which differentiate it from the anchor.

# Review of Previous Verification Test Activities

In this section, MPEG verification test activities of the last 20 years are collected for reference:

* AVC (ISO/IEC 14496-10 | ITU-T Rec. H.264) compared to MPEG-4 Visual (ISO/IEC 14496-2) and MPEG-2 Video (ISO/IEC 13818-2) [4]
  + AVC Baseline and Main profiles
  + QCIF, CIF, SD, HD resolutions
  + Evaluation methods: MM-DSIS (5-grade scale), DSCQS (cont. scale)
* Scalable Video Coding (SVC) extension of ITU-T Rec. H.264 & ISO/IEC 14496-10 Advanced Video Coding (AVC) compared to AVC single layer coding [5]
  + Scalable Baseline, Scalable High, and Scalable High Intra profiles
  + Quality scalability for CIF@30fps video, and spatial scalability for 640x352@60fps video with 1280x704@60fps enhancement
  + Mobile TV with quality scalability for QVGA@25fps video, and spatial scalability for QVGA@12.5fps with VGA@25fps enhancement
  + HD TV with spatial scalability for 720p@50fps with 1080p@50fps enhancement
  + Movie production with spatial scalability for 1080p@25fps being the highest resolution, with two lower resolutions
  + Evaluation methods: SSMM, DSUR (11-grade scale)
* HEVC Main profile compared to the AVC High profile [6]
  + 480p, 720p, 1080p, UHD
  + Evaluation method: DCR (11-grade scale)
* HEVC on 8 bit 4:2:0 interlaced-scan video, 10 bit 4:2:2 progressive-scan video, and 10 bit 4:4:4 progressive-scan video test material compared to AVC [7]
  + HEVC the Main, Main 4:2:2 10, and Main 4:4:4 10 profiles
  + 1080i, 1080p
  + Evaluation method: DCR (11-grade scale) [not documented in report]
* SHVC Scalable Main and Scalable Main 10 profiles compared to HEVC simulcast using the HEVC Main and Main 10 profiles [8]
  + Spatial 2x, spatial 1.5x, SNR, and CGS scalability
  + UHD, HD resolutions
  + Evaluation method: DCR (11-grade scale)
* Supplemental SHVC verification test, comparison to HEVC Main10 Profile [9]
  + SHVC Scalable Main 10 profile compared to HEVC *simulcast* using the HEVC Main 10 profile
  + 1080p resolution, SDR and HDR (BT.2020 PQ)
  + Evaluation method: DCR (11-grade scale)
* Internet Video Coding compared to AVC High Profile [10]
  + Resolutions 832×480 “WVGA”, 1280×720 “720p” and 1920×1080 “HD”
  + Evaluation method: DCR (11-grade scale)
* HEVC Screen Content Coding Extensions (SCM) compared to HEVC (HM) and AVC (JM) [11]
  + TGM (text and graphics with motion) and mixed (video and graphs mixed), and for the three colour subsampling methods considered (RGB 4:4:4, YCrCb 4:2:0 and YCrCb 4:4:4)
  + Evaluation method: DCR (11-grade scale)
* EVC compared to AVC and HEVC [12][13]
  + Main profile compared to the HEVC Main 10 profile
  + Baseline profile compared to the AVC Baseline profile
  + SDR, HDR PQ,
  + UHD and HD EVC resolutions
  + Evaluation method: DSIS/DCR (11-grade scale)
* LCEVC compared to AVC, EVC, HEVC, and VVC [14]
  + HD and UHD resolutions
  + Evaluation method: DCR (11-grade scale)
* VVC compared to HEVC [15][16]
  + Comparison of Main 10 profiles
  + HD and UHD resolutions, 360° video, gaming-type content
  + Random access and low delay configurations
  + Evaluation method: DCR (11-grade scale)

# Conclusions

This document describes the guidelines for verification testing of standards which define visual media output. The guidelines represent the best practice as consented by AG5 and the MPEG WGs. As documented by the list of previous verification tests, it is founded on the experience gained in the former MPEG structure and has been driven by the MPEG Test Chairs.

# References

1. International Telecommunication Union – Radio Communication Sector; “Methodologies for the subjective assessment of the quality of television images.” Recommendation ITU-R BT.500-14, Oct. 2019.
2. International Telecommunication Union – Telecommunication Standardization Sector; “Subjective video quality assessment methods for multimedia applications.” Recommendation ITU-T P.910, Apr. 2008.
3. International Telecommunications Union – Telecommunications Standardization Sector; “Working practices using objective metrics for evaluation of video coding efficiency experiments.” Technical Paper HSTP-VID-WPOM, July 2020. <http://handle.itu.int/11.1002/pub/8160e8da-en>
4. ISO/IEC JTC1/SC29/WG11, “Report of The Formal Verification Tests on AVC (ISO/IEC 14496-10 | ITU-T Rec. H.264),” Doc. N6231, Waikoloa, HI, USA, Dec. 2003
5. ISO/IEC JTC1/SC29/WG11, “SVC Verification Test Report,” Doc. N9577, Antalya, TR, Jan. 2007
6. ISO/IEC JTC1/SC29/WG11, “Report on HEVC compression performance verification testing,” Doc. N14420, Valencia, ES, Apr. 2014
7. ISO/IEC JTC1/SC29/WG11, “HEVC interlaced video and format range extensions verification test results,” Doc. N15437, Warsaw, PL, Jun. 2015
8. ISO/IEC JTC1/SC29/WG11, “SHVC verification test report,” Doc. N16051, San Diego, CA, USA, Feb. 2016
9. ISO/IEC JTC1/SC29/WG11, “Supplemental SHVC verification test report,” Doc. N16268, Geneva, CH, Jun. 2016
10. ISO/IEC JTC1/SC29/WG11, “Verification test report for Internet Video Coding,” Doc. N16255, Geneva, CH, Jun. 2016
11. ISO/IEC JTC1/SC29/WG11, “Verification Test Report for HEVC Screen Content Coding Extensions,” Doc. N16882, Hobart, AU, Apr. 2017
12. ISO/IEC JTC1/SC29/WG04, “Report on Essential Video Coding Compression Performance Verification Testing for HDR/WCG Content,” Doc. 30, online, Oct. 2020
13. ISO/IEC JTC1/SC29/WG04, “Report on Essential Video Coding Compression Performance Verification Testing for SDR Content,” Doc. 47, online, Jan. 2021
14. ISO/IEC JTC1/SC29/WG04, “Verification Test Report on the Compression Performance of Low Complexity Enhancement Video Coding,” Doc. N76, online, Apr. 2021
15. ISO/IEC JTC1/SC29/WG05, “VVC verification test report for UHD SDR video content,” Doc. N21, online, Oct. 2020
16. ISO/IEC JTC1/SC29/WG05, “VVC verification test report for HD SDR and 360-degree video content,” Doc. N54, online, Apr. 2021