|  |  |
| --- | --- |
| **Joint Collaborative Team on 3D Video Coding Extension**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  13th Meeting: Geneva, CH, 17 – 21 Oct 2015 | Document: JCT3V-M1002 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **MV-HEVC Verification Test Plan** | | |
| *Status:* | Output Document | | |
| *Purpose:* | Test plan report | | |
| *Author(s) or Contact(s):* | Karsten Müller, HHI Shinya Shimizu | Email: | [karsten.mueller@hhi.fraunhofer.de](mailto:karsten.mueller@hhi.fraunhofer.de)  [shimizu.shinya.0827@gmail.com](mailto:shimizu.shinya.0827@gmail.com) |
| *Source:* | JCT-3V | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This document is the verification test plan for MV-HEVC, including test conditions, evaluation methodology and timeline.

# Introduction

The primary usage scenario for 3D video is to support applications where 3D depth perception of a visual scene is provided by a 3D display system. There are many types of 3D display systems including classic stereo systems that require special-purpose glasses and those with more sophisticated multiview auto-stereoscopic displays that do not require glasses.

A new generation of 3D Video Coding technology that goes beyond the capabilities of existing standards to enable both advanced stereoscopic display processing and improved support for auto-stereoscopic multi-view displays has been a primary subject of work by the JCT-3V group. A new data format and associated compression technology to enable the high-quality reconstruction of synthesized views for 3D displays have been developed for both AVC and HEVC-based coding frameworks. As part of this work, a multi-view coding extension of HEVC (MV-HEVC) has been developed. The main target of MV-HEVC is to enable coding multi-view video sequences. Depth maps can be coupled with multi-view video streams using auxiliary pictures, which are one of the features in the multi-layer extensions of HEVC. There are no changes to the CU-level syntax, semantics and decoding processes of HEVC. The specification of this extension was provided in the Annex G of HEVC, which was published in ITU-T Rec. H.265 edition 2 (Consented 2014-07, approved 2014-10, published 2015-01-12) and ISO/IEC 23008-2 edition 2 (approval to ballot 2014-07, text submission 2014-10, published 2015-05-01).

As the standardization of the specification text has been completed and adequately stable software has been developed, a verification test is planned to assess the improvement of the coding performance. MV-HEVC will be compared with simulcast coding of HEVC as well as with MVC (the multiview extension of AVC) in terms of stereo video coding. In the following sections, the timeline of these tests, the test conditions and evaluation procedure are described.

# Timeline

2015/06 (12th JCT-3V Meeting): Made test bitstreams and viewing material available

Transferred viewing material to Vittorio Baroncini (FUB)

2015/11 – 2016/02: Perform subjective viewing test

2016/02 (14th JCT-3V Meeting): Provide the subject viewing result as an input document

# Test Conditions

## Input Test Sequences

The multiview test sequences with associated depth data, and corresponding input views to be used for experiments are specified in the table below.

|  |  |  |
| --- | --- | --- |
| **Seq. ID** | **Test Sequence** | **left-right** |
| S03 | Undo\_Dancer | 3-5 |
| S04 | GT\_Fly | 5-3 |
| S13 | Band06 | 0-1 |
| S14 | BMX | 0-1 |

The followings are spare sequences which will be used if clear quality differences among quality points are not evident.

|  |  |  |
| --- | --- | --- |
| **Seq. ID** | **Test Sequence** | **left-right** |
| S11 | Musicians | 0-1 |
| S10 | Shark | 5-6 |

## Encoder Configuration

Three coding frameworks are considered for the subjective test:

* MVC: AVC-based multiview video coding (non-base view is coded using inter-view prediction)
* Simulcast HEVC: each view is coded independently
* MV-HEVC: HEVC-based multiview video coding (non-base view is coded using inter-view prediction)

The encoder configuration settings for both encodings are consistent with the common test conditions (CTC) as given in JCT3V-G1100, which are also outlined below:

* Inter-view coding structure
* 2 view case: left-right (in coding order)
* Temporal prediction structure: GOP 8, intra every 24 frames (random access at ~1sec)
* Full resolution texture coding
* Software: JM v18.6, 3D-HTM v14.1 (to be used in Simulcast and MV-mode)
* Encoder configurations for 3D-HTM are provided as part of the “Starter-kit” (./cfg/MV-HEVC)

Bit rates/QP settings:

* Target 4 rate points
* HEVC Independent view texture QP values: 25, 30, 35, 40
* MV-HEVC Dependent view texture QP values: Started with same QP offset values defined in the CTC and adjusted QP values to approximately match PSNR
* MVC QP values: Found QP values that approximately match PSNR

The table below summarizes the QP settings obtained for each test sequence.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **QP values (Independent view/dependent view)** | | |
| **Seq. ID** | **Test Sequence** | **MVC** | **Simulcast HEVC** | **MV-HEVC** |
| S03 | Undo\_Dancer | 23/25, 28/29, 32/35, 37/39 | 25, 30, 35, 40 | 24/27, 30/33, 35/38, 40/43 |
| S04 | GT\_Fly | 23/24, 27/30, 32/33, 36/37 | 25, 30, 35, 40 | 24/27, 29/32, 35/38, 39/42 |
| S10 | Shark | 23/24, 28/29, 32/33, 36/37 | 25, 30, 35, 40 | 25/28, 30/33, 35/38, 40/43 |
| S11 | Musicians | 23/24, 27/29, 32/33, 36/39 | 25, 30, 35, 40 | 24/27, 29/32, 34/37, 39/42 |
| S12 | Poker | 22/25, 27/28, 31/33, 35/38 | 25, 30, 35, 40 | 24/27, 29/32, 34/37, 39/42 |
| S13 | Band06 | 23/25, 28/29, 32/34, 36/39 | 25, 30, 35, 40 | 24/27, 29/32, 34/37, 39/42 |
| S14 | BMX | 22/24, 26/28, 30/32, 34/37 | 25, 30, 35, 40 | 24/27, 29/32, 34/37, 39/42 |

# Expert Viewing Protocol

The visual evaluation provided by the three coding schemes will be made by means of an “Expert Viewing Protocol” (EVP), as used for two previous MPEG subjective evaluation experiments [2] [3].

The EVP is based on the participation of MPEG experts, who were not directly involved in the activities related to the tested video materials.

The EVP is a variation of the “Double-Stimulus Impairment Scale” (DSIS) test method (as described by the ITU-R Recommendation BT-500 [1]) where the modifications introduced are:

1. Only 9 experts participate as viewers in each EVP session,
2. The “unimpaired” Source video Clip (SRC) is shown once, every two Processed Video Clips (PVS).

Therefore, the viewing timing of an EVP Basic Test Cell (BTC) is set up as shown in *Figure 1*.

*“BTC N” SRC video “A” PVS video A “B” PVS video B “Vote A and B”*

*(1 sec.) (10 sec.) (1 sec.) (10 sec.) (1 sec.) (10 sec.) (5 sec.)*

*Figure 1 – Time scale of EVP Basic Test Cell*

Here, the captions “BTC N” “A” “B” Vote A and B” represent messages that are displayed on the screen.

This BTC timing allows to save a considerable amount of time in relation to a standard DSIS test protocol, since the experts have more habit in viewing images and can remember much better the details of a reference video clip.

The time required by EVP to evaluate two PVSs is 38 seconds, against a total of 54 seconds required by the DSIS method, thus saving approximately 30% of viewing time.

Furthermore an EVP does not requires a stabilization phase as well as the insertion of one or more SRC vs SRC test cells, dueto the high ability of experts to create their own evaluation scale and to be reliable in a way that no SRC vs. SRC check is required. Thus, the overall session length can further be reduced by more than 15%.

## Viewing area set-up

A 3D display of diagonal size greater than or equal to 40” will be used.

Three viewers will be seated in front of the HD 3D monitor (or high quality TV set), at a distance of 3 H (3 times the screen height), taking care that, in any case, the widest viewing angle does not exceed 60° from the center axis of the screen. The position of viewers has to be recorded, to allow a post experiment verification of the influence of the viewing position (i.e. center, left, right) on the MOS value.

The testing area has to be completely dark and any visible and audible pollution has to be avoided.

A low power (e.g. 25 Watts or less) light source is placed behind the monitor and directed to the wall behind the monitor in a way, that no direct light points to the viewers. The distance between display and viewers shall be ≥ 1m in order to avoid distraction of the viewers.

The light behind the monitor has two functions: to allow the viewers to see their scoring sheets and to mitigate any sudden light changes on the monitor.

## Test design

The test will consist of as many test sessions as necessary to evaluate all the PVSs.

The orders of presentation of the PVSs inside a BTC are randomly changed (to avoid any bias in the judgment) in a way hidden to the viewers.

A 10-grade impairment scale is used to assess the visual quality of the coded video clips.

### 10 grades impairment scale and viewers’ training

Even if the viewers are all “experts” in the area of video processing, they need a short training about:

* Timing of presentation of the video clips on the screen,
* How to fill out the scoring sheet,
* Meaning and use of the 10 grade impairment scale.

The 10-grade impairment scale allows viewers to express a judgement of the degradation (if any) between the “SRC” and the processed video clips (PVS).

A short training session (namely 6 BTCs) is run to let the viewers understand when to look at the screen and when to look at the scoring sheet, and how and when to express their opinion.

The BTCs of the training session must include PVSs equally representing the overall impairment range of a test session.

The viewers are explained to carefully look at the video clips shown immediately after the message “A” and “B”, to notice if they were able to see any difference with the video clip shown after the message “BTC N”.

The following guidance about the meaning of the numerical scoring will be given to the viewers:

* should the viewer not be able to see any difference between the source and video clip “A”, a score of 10 is written in box “1” of a BTC (see Figure 2). Similar, if no difference between the source and video clip “B” occurs, a score of 10 is written in box “2”.
* in the case any, even very small, impairments are visible, a score of “9” is given, if the impairment is just in one area of the image, or “8” if the difference is noted in many areas of the screen.
* Scores 7 and 6 are given, if the impairment are clearly visible.
* Scores 5 and 4 are given, if impairments are evident at first sight.
* Scores 3 and 2 are given, when impairments are annoying.
* Scores 1 and 0 are given, when the image is severely corrupted, in some area or everywhere.

### Scoring Sheet

An example of the scoring sheet for an EVP session is shown below.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | BTC 1 | | | | |  | BTC 2 | | | | |  |  |  |  | BTC n/2 | | | | |  |
|  |  | 1 |  | 2 |  |  |  | 3 |  | 4 |  |  | … | |  |  | n-1 |  | n |  |  | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |

*Figure 2 – Scoring Sheet*

## Data analysis

At least nine subjects must participate in each session. This means that, when more than one session has to be run to complete the evaluation, each session must be run with nine viewers, however it is not mandatory that the same viewers run all the sessions.

The obtained viewing results will be statistically analyzed, computing the mean-opinion score (MOS) and the confidence interval (CI) for each test. It is noted that with nine scores the computation of the CI already provides a good indication of when two coding condition are assumed to be different in visual quality.

## Usability and stability of the expert viewing procedure

In other EVP tests ([2][3]) the ranking of the video clips has been excellent and stable, providing a very good discrimination of the different qualities both in term of relative and absolute values. This encourages the use of EVP also for this Verification test.

1. International Telecommunication Union – Radio Communication Sector; Recommendation ITU-R BT.500-13.
2. WG11, “Results of Call for Evidence on High-Performance Video Coding (HVC)”, ISO/IEC JTC1/SC29/WG11 N10721, London, UK, July 2009.
3. WG11, “Report of IVC visual quality evaluation” ISO/IEC JTC1/SC29/WG11 N14989, Strasbourg, FR, October 2014.

Appendix A: Test Sequences

**S01: Poznan\_Hall2**

**S02: Poznan\_Street**

**Poznan University of Technology**

<ftp://multimedia.edu.pl/3DV/>  
username: 3DV

password: ftvftv

directory: CFP

*Copyright: Individuals and organizations extracting sequences from this archive agree that the sequences and all intellectual property rights therein remain the property of the respective owners listed below. These materials may only be used for the purpose of developing, testing and promulgating technology standards and for academic usage. Acknowledgement and reference to the following source document: Marek Domañski, Tomasz Grajek, Krzysztof Klimaszewski, Maciej Kurc, Olgierd Stankiewicz, Jakub Stankowski, Krzysztof Wegner "Poznañ Multiview Video Test Sequences and Camera Parameters", ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, Xian, China, October 2009, are required in all documents that report any usage of the materials. The respective owners make no warranties with respect to the materials and expressly disclaim any warranties regarding their fitness for any purpose.*

Owners: Poznan University of Technology, Poznañ, Poland.

**S03: Undo\_Dancer**

**S04: GT\_Fly**

**Nokia**

<ftp://mpeg3dv.research.nokia.com>

username: mpegmember

password: S9"12#sHD)3

*Copyright: The supplied data and content (“Supplied Data”) is provided free of charge and made available for use by the Licensee, who shall be a member of the MPEG standardization committee or a respondent to a the Call for Proposals for the MPEG standard for 3D video coding that is to be issued March 2011, under the following conditions:*

*The Licensee agrees that the Supplied Data and all intellectual property rights therein remain the property of Nokia Corporation and its licensors (owners). The Supplied Data may only be used for the purpose of responding to the Call for Proposals that is to be issued March 2011, developing, testing and promulgating the MPEG technology standard for 3D video coding resulting from the Call for Proposals that is to be issued March 2011. The respective owners make no warranties with respect to the Supplied Data and expressly disclaim any warranties regarding its fitness for any purpose.*

*As a way of promulgating the MPEG technology standard for 3D video coding, the Licensee may present parts or modifications of the Supplied Data at academic conferences and publications.*

*The Licensee agrees not to provide the data to any third parties without permission from the owners and that the data shall not be sold, let for hire, or by way of trade, offered or exposed for sale or hire. This restriction shall apply to the original material or to any reproduction of it in whole or in part and to any modifications of the Supplied Data.*

*The texture views of the Supplied Data are copyright © UNDO. (See* [*http://www.undo.fi*](http://www.undo.fi)*)*

Owners: Nokia, Finland.

**S05: Kendo**

**S06: Balloons**

**Nagoya University**

<http://www.tanimoto.nuee.nagoya-u.ac.jp/~mpegftv/mpeg3dv/CfP/>

username: mpegftv

password: fngOyfTv

*Copyright: Individuals and organizations extracting sequences from this archive agree that the sequences and all intellectual property rights therein remain the property of the respective owners listed below. These materials may only be used for the purpose of developing, testing and promulgating technology standards. The respective owners make no warranties with respect to the materials and expressly disclaim any warranties regarding their fitness for any purpose.*

Owners: Nagoya University, Japan.

## S08: Newspaper

**GIST**

<ftp://203.253.128.142>

username: 3DV

password: 3dvkr

directory: /GIST\_Test\_Sequence/Newspaper

*Copyright: Individuals and organizations extracting sequences from this archive agree that the sequences and all intellectual property rights therein remain the property of the respective owners listed below. These materials may only be used for the purpose of developing, testing and promulgating technology standards. The respective owners make no warranties with respect to the materials and expressly disclaim any warranties regarding their fitness for any purpose.*

Owners: Gwangju Institute of Science and Technology (GIST), Republic of Korea.

## S10: Shark

**NICT**

ftp://ftp.merl.com

username: anonymous

password: <your e-mail address>

directory: /pub/tian/NICT-3D/Shark

*Copyright: The supplied data and content (“Supplied Data”) is provided free of charge and made available for use by the Licensee, who shall be a member of the JCT-3V standardization committee, under the following conditions:*

*The Licensee agrees that the Supplied Data and all intellectual property rights therein remain the property of the National Institute of Information and Communications Technology (NICT) and its licensors (owners). The Supplied Data may only be used for the purpose of developing, testing and promulgating the JCT-3V technology standards and for academic usage. The respective owners make no warranties with respect to the Supplied Data and expressly disclaim any warranties regarding its fitness for any purpose.*

*As a way of promulgating the technology standards, the Licensee may present parts or modifications of the Supplied Data at academic conferences and publications.*

*In the use of the parts or modifications of the Supplied Data, the Licensee shall make the source clear in appropriate manners.*

*The Licensee agrees not to provide the data to any third parties without permission from the owners and that the data shall not be sold, let for hire, or by way of trade, offered or exposed for sale or hire. This restriction shall apply to the original material or to any reproduction of it in whole or in part and to any modifications of the Supplied Data.*

Owners: National Institute of Information and Communication Technology (NICT), Japan