**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2018/N17467**

**January 2018, Gwangju, Korea**

|  |  |
| --- | --- |
| **Source** | Video |
| **Status** | Approved |
| **Title** | Common Test Conditions on 3DoF+ Visual |
| **Editors** | Bart Kroon, Miska M. Hannuksela, Renaud Doré, Mary-Luc Champel, Jill Boyce |

# Introduction

The common test conditions are defined to evaluate the coding efficiency and subjective quality and user experience of solutions supporting head motion parallax effect for omnidirectional video.

Related to this document are:

* [N17466](http://wg11.sc29.org/doc_end_user/current_document.php?id=61340&id_meeting=) *Investigation of 3DoF+ Visual*, and
* [N17471](http://wg11.sc29.org/doc_end_user/current_document.php?id=61345&id_meeting=) *Call for Test Materials for 3DoF+ Visual*.

# Test procedures

To progress the technical work in a systematic manner, the following steps are proposed:

1. Compress test content with existing codec technology, e.g. HEVC/MV-HEVC/3D-HEVC,
2. Synthesize virtual views by decoded views and metadata,
3. Render view ports of real/virtual pose traces of limited head movement, typically sitting in a seat,
4. Evaluate coding efficiency and parallax effect.

The stream shall be viewer independent, meaning that neither the position nor the orientation of the viewer shall be taken into account when compressing the test content. The range of supported possible viewer positions will be constrained and known, as defined in Section 5.2.

## Decisions on test procedure

At the next MPEG meeting, the group has to take decisions on common test conditions based on new information. In the remainder of this document, we will assume that all decisions are positive:

* At least one intermediate view synthesizer will be submitted in response to this document. Based on the availability, quality and documentation of this tool, the group will accept it as the *Reference Intermediate View Synthesizer* (RIVS, Section 4.2).
* Content will be submitted in response to the call for test materials ([N17471](http://wg11.sc29.org/doc_end_user/current_document.php?id=61345&id_meeting=)) and sufficient sequences will be accepted,
* Head-mounted display (HMD) pose traces (Section 3.2) will be submitted in response to this document, and they will be accepted.
* Anchor configurations will be defined per sequence, filling in the tables in this document.
* A detailed timeline for 3DoF+ investigations should be defined.

The general evaluation process according to above decision points is depicted in Figure 1. For the anchor there are additional evaluation steps to assess the performance of compression and view synthesis (Figure 2).

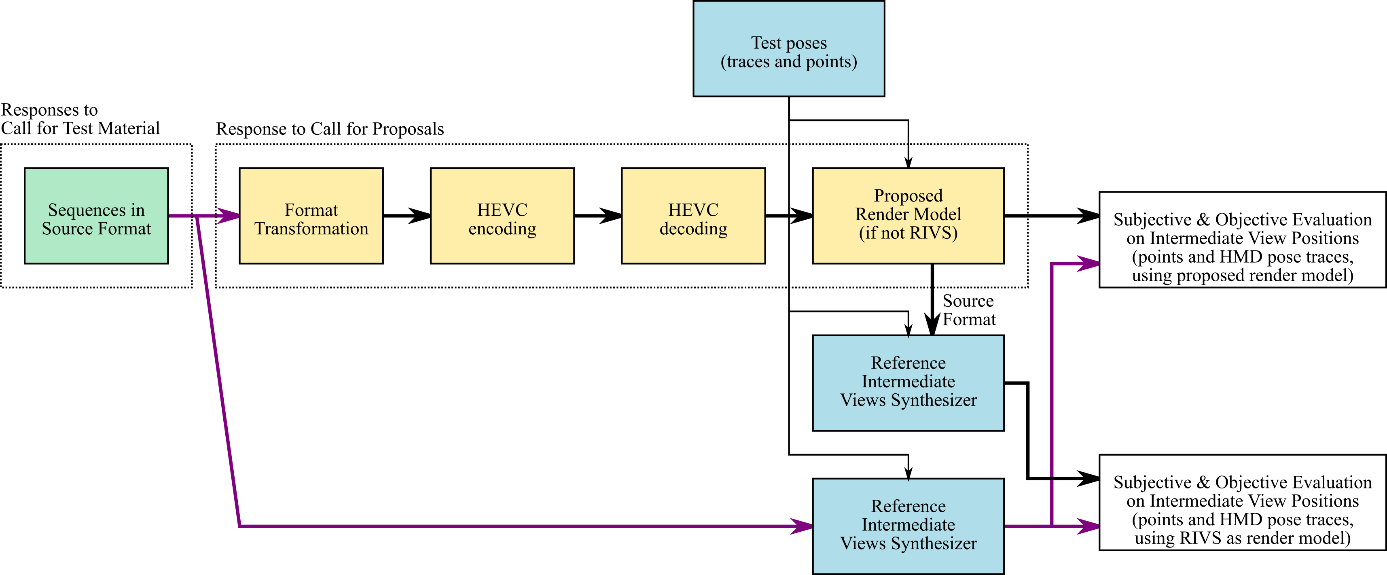


Figure 1: The general evaluation process (for anchor and proposals)

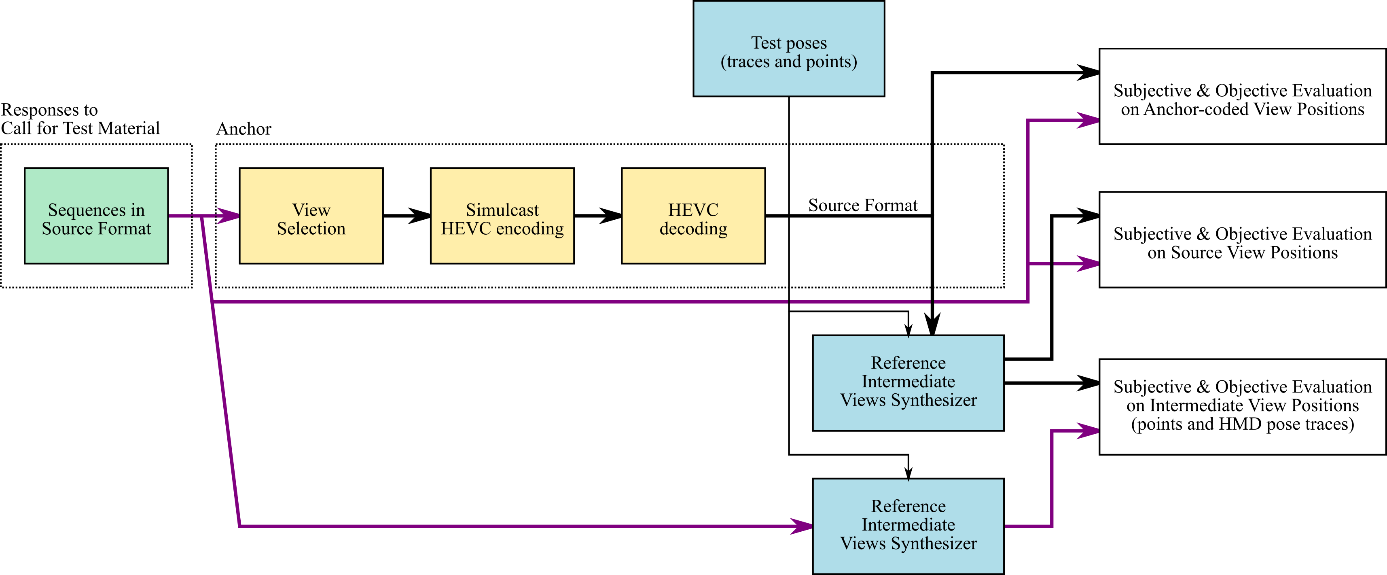


Figure 2: The extension of the evaluation process for the anchor

# Test materials

## Video files

Test materials will be provided according to the source format described in [N17471](http://wg11.sc29.org/doc_end_user/current_document.php?id=61345&id_meeting=). The original data (e.g. 3D models or camera acquisitions) may remain internal to the participants. Even multi-view, single-frame sequences (i.e. multi-view images) are welcome both for objective and subjective evaluation.

## HMD pose traces and points

To evaluate the performance of a technical solution, multiple files will be made available that contain a position and translation (in meters) for each frame. We hereby define such data as a *HMD* *pose trace*. Such a trace can be generated (as preparation, and not as part of a technical solution) for instance by inviting human test participants to wear an HMD and behave according to the 3DoF+ application: sitting with natural head motion or standing, but not taking any steps. Such data can be provided already into the next meeting for internal testing[[1]](#footnote-1).

# Software Tools

## Codec

Both anchor and CfP responses shall use HEVC Main 10 profile for the compression of video data. Hereby the total pixel rate shall not be greater than in HEVC Level 6.2 (i.e. 8K@120Hz), and everything else it is up to proponents (e.g. framepacking or multicast). Proponents are allowed to apply transformations to the source format before encoding and after decoding. Examples are view selection, packing and projection.

## Reference Intermediate views synthesizer (RIVS)

The RIVS[[2]](#footnote-2) is capable of generating in an offline fashion both:

1. A rectilinear rectangular viewport based on a set of camera parameters (e.g. rotation, translation, resolution).
   * The frames will be 2048 x 2048, 90-degree field of view, and the same bit depth as the input material.
2. An ERP frame of the whole 360 content, for a specified position (x’, y’, z’) in meters.
   * The frames will be the same resolution (4K or higher) and bit depth (8 or 10 bit) as the input material.

The RIVS is part of the anchor (Figure 1) and should be used by proponents as well (Figure 2).

## Render model

A solution requires a render model similar to RIVS that is capable of providing rectangular viewports and ERP frames. The render model shall accept the same pose parameters as provided to generate the reference views. However, this render model takes as input the content in the format as output by a proposed decoder (Figure 1).

The anchor will use RIVS as the render model and a proponent may choose RIVS as well. When a proponent chooses to provide a render model as part of the proposed technical solution, then the proponent should evaluate the performance difference between RIVS and the proposed render model by converting the output back to the source format. Valid reasons for a proponent to ignore this rule are that the conversion from the proposed format to the source format would take too much effort or that it would degrade quality. The proponent shall provide a motivation, because by using the RIVS before and after encoding, evaluation concentrates on the format and excludes rendering aspects.

NOTE: While test evaluations require a common software renderer, in NO way shall such renderer be normative in OMAF.

# Test conditions

## Anchor

The anchor will be simulcast HEVC encoding of a subset of the source views, which are referred to as the anchor coded views, as indicated in the table below. For some sequences, the subset may be the complete set.

Table 1: Anchor coded views

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Name | # of source views | # of anchor coded views | anchor coded views |
| A | Sequence1 | 7 | 3 | 1, 3, 5 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

The common test conditions defined in Section 6 will be used. To calculate intermediate views, the RIVS software will be used, with the anchor coded views as inputs.

## Contributions

Proposals are not required to code views corresponding to all of the anchor coded views, but are required to be able to generate viewport and ERP video sequences for any intermediate view position in the designated range for each test sequence. The field of view for ERP (e.g. 180° or 360° degrees) will be the same as the source content.

Table 2: Intermediate view position ranges

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Name | x\_min | x\_max | y\_min | y\_max | z\_min | z\_max |
| A | Sequence1 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

# Evaluation procedures

## Objective evaluation

### Proposals

For all ERP sequences and both for anchors and proposals, WS-PSNR based BD-rate values will be provided for specific intermediate view positions (Table 3).

Table 3: Intermediate view positions for objective evaluation (per sequence)

|  |  |  |
| --- | --- | --- |
| Class | Sequence name | Location of intermediate tested views |
| A | Sequence1 |  |
|  |  |  |
|  |  |  |

### Anchor

In addition, only the anchor is evaluated on anchor coded views without applying RIVS and non-coded source positions (Table 1). Purpose of this additional evaluation is to study the quality of compression and view synthesis in separation:

1. Anchor coded and source view positions: WS-PSNR BD-rate provided for each source view position in ERP, presented individually for each view, combined for all anchor coded views, and separately combined for all source views.
2. Intermediate view positions:WS-PSNR BD-rate provided for each of the pre-determined intermediate view positions in ERP, as indicated in the table above.

Templates for reporting test results for HM and JEM coding can be found at link. [template should be provided which contains anchor results.]

## Subjective evaluation

Contributions may be subjectively evaluated for four categories of viewpoints

1. Anchor coded view position
2. Source view position
3. Static intermediate view position
4. Dynamic intermediate view position

The view position and orientation of the source and intermediate views may be pre-defined, or may be defined later. For dynamic intermediate views, HMD pose traces shall be used.

Table : HMD pose traces/sets for subjective evaluation (per sequence)

|  |  |  |
| --- | --- | --- |
| Class | Sequence name | HMD pose trace/set file |
| A | Sequence1 |  |
|  |  |  |
|  |  |  |

For the anchor, the intermediate view will be generated at the particular intermediate view position for each tested sequence using the RIVS software to generate a rectilinear view with a field of view and resolution as specified in Section 4.2.

For contributions, both RIVS and the render model (when different) is used to generate a rectilinear view with the same parameters as the anchor.

## Test devices

The render model is offline and the resulting videos will be watched:

* Viewport test videos will be watched on a 2D monitor, and
* ERP test videos will be watched on a head mounted display.

Equipment will be arranged as defined in the CfP. Suitable equipment may be a mobile workstation with sufficient main memory to cache the longest test video (e.g. 16 or 32 GB) and an Oculus Rift headset.

# Encoder configurations and quantization parameter values

The test conditions will be aligned with the JVET common test conditions for the HM anchor, using the random access 10-bit case. For each video sequence, four quantization parameter values[[3]](#footnote-3) are to be used: 22, 27, 32 and 37.

To be aligned with JVET and JCT-VC CTC, motion search range is set to 256, and GOP size is set to 16.

1. (At least) Philips will provide an input contribution with pose traces that match the 3DoF+ viewing zone description. [↑](#footnote-ref-1)
2. Technicolor may be willing to provide such software [↑](#footnote-ref-2)
3. Fixed QPs should be defined for CTCs. In a CfP, target bitrates would instead be used. [↑](#footnote-ref-3)