|  |  |
| --- | --- |
| **Joint Collaborative Team on 3D Video Coding Extension Development**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  2nd Meeting: Shanghai, CN, 13–19 Oct. 2012 | Document: JCT3V-B1110 |

|  |  |  |
| --- | --- | --- |
| *Title:* | **Common Test Conditions of 3DV Core Experiments** | |
| *Status:* | Output Document | |
| *Purpose:* | Common Test Conditions | |
| *Author(s)):* | Dmytro Rusanovskyy  Karsten Müller Anthony Vetro | [Dmytro.Rusanovskyy@nokia.com](mailto:Dmytro.Rusanovskyy@nokia.com) [karsten.mueller@hhi.fraunhofer.de](mailto:karsten.mueller@hhi.fraunhofer.de) [avetro@merl.com](mailto:avetro@merl.com) | |
| *Source:* | JCT-3V | |

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

This document specifies the common test conditions for 3D video core experiments with AVC-based and HEVC-based test models.

# Test Cases

There are two general test scenarios under consideration:

* multiview/stereo video coding with depth data
* multiview/stereo video coding without depth data

For each test scenario, experiments consider a two-view case (C2) and a three-view case (C3), where C2 is subset of C3 coding configuration. Simulations are run under the C3 configuration and the corresponding results are reported for C3. The C2 results are derived from the C3 results using the recommended Excel template.

Proponents are encouraged to directly simulate the C2 configuration if they find it useful for their specific proposal.

# Test Sequences

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Seq. ID** | **Test Sequence** | **2-view input** | **3-view input** |
| S01 | Poznan\_Hall2 | 7-6 | 7-6-5 |
| S02 | Poznan\_Street | 5-4 | 5-4-3 |
| S03 | Undo\_Dancer | 1-5 | 1-5-9 |
| S04 | GT\_Fly | 9-5 | 9-5-1 |
| S05 | Kendo | 1-3 | 1-3-5 |
| S06 | Balloons | 1-3 | 1-3-5 |
| S08 | Newspaper1 | 2-4 | 2-4-6 |

# Basic Encoder Configuration

The encoder configuration settings that are common to both HEVC- and AVC-based experiments are given below:

* Inter-view coding structure
  + 3 view case: center-left-right (in coding order)
  + 2 view case: right-left (in coding order)
  + P-I-P inter-view prediction
* Temporal prediction structure
  + GOP 8
  + intra every 24 frames (random access at roughly each second)

Specific encoder configurations for AVC-based experiments are specified as follows:

* Anchor software: 3DV-ATM v6.0
* <http://mpeg3dv.research.nokia.com/svn/mpeg3dv/tags/>
* Encoder HP/EHP configurations are provided as part of the “Starter-kit”
  + ./configs
* Changes to encoder configurations compare to previous configuration:
  + EHP profile:
    - ALC (B0031) ON
* Independent view texture QP values: 26, 31, 36, 41 (Full resolution)
* Independent view depth QP values: 26, 31, 36, 41 (Reduced resolution)
* Full resolution texture coding
  + Check sum of input data to be provided in the “Starter-kit”
* Reduced resolution depth coding (50% in each direction)
  + Downsampling: JSVM utility, method1 (included in the Starter-kit)
  + Up-sampling: built-in 3DV-ATM (bilinear).
  + Depth map post-processing (JCT2-A0038/m25885)
  + Check sum of the input data to be provided in the “Starter-kit”

Specific encoder configurations for HEVC-based experiments are specified as follows:

* Anchor software: 3DV-HTM 5.0
  + <https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware>
* Encoder configurations are provided as part of the “Starter-kit”
  + <https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/configsEncoder>
  + Configurations identical to configuration of 3DV-HTM
* Changes to encoding configuration compare to the previous version:
  + VSO with depth fidelity term ON (B0131)  
    with weight factors 10:1:1 for VSO:SAIT\_RDO:DEPTH
  + QT Prediction ON (B0068)
* 8-bit input data to be used
* HEVC codec is configured with 8-bit internal processing
* Encoder configuration files and renderer configuration provided in the Starter-kit
* Full resolution of video and depth
* Texture QP values for independent view: 40, 35, 30, 25
* Depth QP values: fixed relation to video QP as shown below

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| QPV0 | 51 | 50 | 49 | 48 | 47 | 46 | 45 | 44 | 43 | 42 | 41 | 40 | 39 | 38 | 37 | 36 | 35 | 34 | 33 | 32 | 31 | 30 | 29 | 28 | 27 | 26 | 25 |
| QPD0 | 51 | 50 | 50 | 50 | 50 | 49 | 48 | 47 | 47 | 46 | 45 | 45 | 44 | 44 | 43 | 43 | 42 | 42 | 41 | 41 | 40 | 39 | 38 | 37 | 36 | 35 | 34 |

# Objective evaluation of video quality

* Reporting template is provided in the “Starter-kit” package
* View synthesis distortion (PSNR) computed against reference view synthesis results (R0)
* Reference view synthesis (R0) is produced with original texture and depth map data, all at full resolution, original camera parameters information
* Check sums for R0 yuv files to be provided with “Starter-Kit”.
* Use 3 intermediate views between two cameras (evenly distributed), configuration for VS to be provided in the “Starter-kit”.
* View synthesis uses coded camera parameters information.
* Bjontegard deltas calculated from average PSNR and bitrate for each MVD component of the coded views independently
* Bjontegard deltas calculated from average PSNR for the synthesized views and the overall bit rate (texture, depth and supplementary data)
* Bjontegard deltas calculated from average PSNR for coded texture and the overall bit rate (texture, depth and supplementary data)

# Subjective evaluation of visual quality

Subjective evaluation of proposed techniques relative to anchors generated according to the CTC may be performed with views specified in the table below, unless otherwise specified in the Core Experiment description.

|  |  |  |  |
| --- | --- | --- | --- |
| **Seq. ID** | **Test Sequence** | **Input views** | **Synthesized**  **Stereo pair** |
| S01 | Poznan\_Hall2 | 7-6-5 | (6.250-5.750) |
| S02 | Poznan\_Street | 5-4-3 | (4.250-3.750) |
| S03 | Undo\_Dancer | 1-5-9 | (4.0-6.0) |
| S04 | GT\_Fly | 9-5-1 | (6.0-4.0) |
| S05 | Kendo | 1-3-5 | (2.5-3.5) |
| S06 | Balloons | 1-3-5 | (2.5-3.5) |
| S08 | Newspaper1 | 2-4-6 | (3.5-4.5) |

# Data to be provided

The following data must be provided as part of the CE report:

* rate-distortion performance relative to the anchor
  + multiview coding without depth:
    - PSNR for each view
    - rate for each view
  + multiview coding with depth:
    - rate for each coded view of each MVD component
    - rate for any additional data, e.g. camera params (if applied)
    - PSNR for coded views
    - PSNR for intermediate views (see objective measurement)
    - BD rates as defined above(see objective measurement)
* Encoder and decoder run times in comparison to anchor
* Decoder run time is collected with disabled writing YUV data to the disk
* If new rendering is tested: renderer run time in comparison to anchor

# View Synthesis Algorithm

The view synthesis algorithm to be used for both AVC and HEVC experiments is the “VSRS-1D-Fast” software based on the version used in core experiments for HTM development. (<https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware>)

The view synthesis used to generate reference results (R0) shall be based on the HTM rendering software using version 3.0 or later.

# Appendix A: Configuration Settings for View Extrapolation

This section describe configuration settings intended for experiments on view extrapolation, which are provided for reference and are not part of the CTC for all experiments.

In the extrapolation experiments, the input views and output view positions that should be synthesized are specified in the table below.

* Interpolated and Extrapolated View

|  |  |  |  |
| --- | --- | --- | --- |
| **Seq. ID** | **Test Sequence** | **Input views** | **Stereo pairs** |
| S01 | Poznan\_Hall2 | 7-6 | (6.25-5.75) |
| S02 | Poznan\_Street | 5-4 | (4.25-3.75) |
| S03 | Undo\_Dancer | 1-5 | (4.0-6.0) |
| S04 | GT\_Fly | 9-5 | (6.0- 4.0) |
| S05 | Kendo | 1-3 | (2.50-3.50) |
| S06 | Balloons | 1-3 | (2.50-3.50) |
| S08 | Newspaper1 | 2-4 | (3.50-4.50) |

* Extrapolated View

|  |  |  |  |
| --- | --- | --- | --- |
| **Seq. ID** | **Test Sequence** | **Input views** | **Stereo pairs** |
| S01 | Poznan\_Hall2 | 7-6 | (6.0-5.5), (5.875-5.375) |
| S02 | Poznan\_Street | 5-4 | (4.0-3.5), (3.875-3.375) |
| S03 | Undo\_Dancer | 1-5 | (5.0-7.0), (5.5-7.5) |
| S04 | GT\_Fly | 9-5 | (5.0- 3.0), (4.5-2.5) |
| S05 | Kendo | 1-3 | (3.0-4.0), (3.25-4.25) |
| S06 | Balloons | 1-3 | (3.0-4.0), (3.25-4.25) |
| S08 | Newspaper1 | 2-4 | (4.0-5.0), (4.25-5.25) |

The VSRS-1D-Fast software shall be used to generate the anchor views for comparison.

View synthesis should be carried out from coded and uncompressed data and the configuration files for generating the anchors, including the view synthesis software (VSRS-1D-Fast) as well as the 3D-ATM or 3D-HTM software, are provided as part of the “Starter-kit” package.