**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2016/N16523**

**October 2016, Chengdu, China**

|  |  |
| --- | --- |
| **Source** | Requirements |
| **Status** | Approved |
| **Title** | Revised Summary of Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation |
| **Editors** | Olgierd Stankiewicz, Krzysztof Wegner, Takanori Senoh, Gauthier Lafruit, Vittorio Baroncini, Masayuki Tanimoto |

# Introduction

The current MPEG state-of-the-art standardized video coding technologies, e.g. HEVC, allow for compression of high quality visual media with resolution even as high as 4k/8k UHDTV (Ultra High-Definition TV). HEVC transmits only a single view and thus the users are viewing 2D content and cannot interactively change the viewpoint. It is still far away from our viewing experience in the real world.

Because of the abovementioned limitations of 2D video codecs, in parallel, MPEG has been developing video coding standards enabling some forms of 3D vision. Previous stereo and multiview coding standards, such as Multi-View Coding (MVC) and MV-HEVC, have focused on the compression of camera views “as is”, without means to facilitate the generation of additional views.

Depth-based 3D formats and associated compression technology (3D-HEVC) have been developed to address this shortcoming: with the use of depth image based rendering techniques, the generation of additional views from a small number of coded views was enabled, supporting auto-stereoscopic 3D display applications with tens of output views from a couple of input camera feeds. 3D-HEVC standard explicitly assume a linear and narrow baseline arrangement of camera inputs.

Recently, more advanced 3D displays – called Super-Multi-View (SMV) displays – are emerging, which render hundreds of linearly or angularly arranged, horizontal parallax ultra-dense views, thereby providing very pleasant glasses-free 3D viewing experience with wide viewing angle, smooth transition between adjacent views, and some “walk-around feeling” on foreground objects. Such displays require both a very dense and a very wide baseline set of views which results in vast number of input views that need to be transmitted. Moreover transmitted and displayed views are highly correlated, which was observed in previously considered applications. As such, it puts very demanding requirements to the codec for both compression efficiency and throughput in order to handle hundreds of wide baseline, ultra-dense views.

The next challenge for MPEG is FTV (Free-viewpoint TV) [1,2,3,4]. FTV enables users to view a scene by freely changing the viewpoints as we do naturally in the real world. FTV is the ultimate 3D-TV experience with an infinite number of views and ranks as the top of visual media. It provides a very realistic glasses-free 3D viewing without eye fatigue. FTV will have a great impact on various fields of our life and society.

FTV covers SMV and Free Navigation (FN) applications [5]. However, SMV displays need huge amounts of data. If compressed with the use of currently available inter-view prediction techniques, the data rate increases linearly with the number of views, beyond economically viable transmission bandwidths.

Literally surrounding the scene with a non-linear, ultra-dense array of several hundreds of cameras offers FN functionalities around the scene, similar to the Matrix bullet effect, further extending the aforementioned “walk-around feeling” to the full scene. Of course, sparse camera arrangements in large baseline setup conditions would be preferable (drastic cost reductions in content generation as well as transmission), however, smooth transitions need to be synthesized. Therefore a key challenge resides in the development of novel technology that supports efficient compression and generation of additional views at the decoder side, which are not already present in the encoded bitstream. Rendering of zoomed-in/out virtual views to support real “fly through the scene” functionalities at reasonable capturing and transmission costs is also required.

Evidently, such 2D rendered FN applications might be combined with SMV for full-immersive viewing on 3D displays. In this case, each FN virtual viewpoint request will synthesize hundreds of linear or angularly adjacent viewpoints to feed the SMV display. This will probably require processing acceleration to cover real-time scenarios, but hardly imposes new algorithmic challenges compared to FN.

With the intention of gathering technologies required for realisation of all aforementioned FN and SMV systems, on 112th meeting in Warsaw, MPEG has issued “**Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation**” (CfE) [6]. It invited companies and organizations that have developed such compression technologies to respond with their proposals. The evaluation of the responses to the CfE was finalized on 115th meeting in Geneva. The results of the CfE have been briefly described in documents [7] and [8] which were outputted from Geneva and San Diego meetings.

This document provides a summary of the CfE, providing description of the submitted technologies, the evaluation methodology and the attained results. All of this is performed in the context of two investigated target application scenarios: Super-Multiview (SMV) displays and Free Navigation (FN).

Though there exist commonalities between SMV and FN, these two categories were evaluated in a different way: SMV aims at high compression exploiting the essential information embedded in all camera views, while improved view synthesis is an additional cornerstone for FN in large baseline arbitrary camera arrangements.

# Application Scenario: Super-Multiview Display (SMV)

## Overview

The source in SMV applications is a large number of views as required by the display. The number of views is typically 80 views or more, and they are arranged in a dense 1D array (linear or arc), covering a wide view range.



Fig. 1. Super-Multiview (SMV) display application scenario.

Compression of a such large amount of captured data requires high bandwidths for transmission. The challenge in SMV consists in improving the coding efficiency to such level that the total bitrate for transmitting hundreds of views stays within realistic and economically viable bandwidth requirements (e.g. data rates corresponding to the transmission of tens of views with today’s technology). This may require the development of new coding paradigms with better view prediction and view synthesis techniques.

## Objective

The main objective in this application scenario is to substantially reduce the data rate required to reconstruct the full set of input views at the receiver compared to existing MPEG state-of-the-art compression standards. The codec may directly transmit all of the input views, use intermediate representations like previously developed multiview plus depth representation with depth estimation and view synthesis embedded in the codec, or may use any other representation which leads to the recreation of the full set of input views at the receiver. At the decoder it is required to recreate the full set of input views.

## Methodology

### Test Material

Only one response to the CfE in SMV category has been received [9][10]. This response was partial and included results for only two sequences out of four sequences in the test set: Big Buck Bunny flowers and Big Buck Bunny butterfly. Description of the used test sequences, along with the exact positions of the views (view numbers), considered as input and output of the codec, is shown in Table 1. The number of views needed to be transmitted is 80 and they are arranged in a dense 1D arc array (Fig. 1). Both sequences are five-second long. For subjective evaluation the sequences have extended twice to last 10 seconds in a ping-pong loop.

Table 1. Summary of the used sequence.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Source | Seq. Name | Number of Views | Resolution (pel) | Frame rate (fps) | Length | Cam Arrangement (1D parall, 1D arc, 2D parall, 2D arc, Sphere, Arbitrary) | Transmitted views positions | Transmitted frame range |
| Holografika | Big Buck Bunny flowers | 91 | 1280x768 | 24 | 5 sec  121 frames | 45 degree arc convergent | 5-84 | 0-120 |
| Holografika | Big Buck Bunny butterfly | 91 | 1280x768 | 24 | 5 sec  120 frames | 45 degree arc convergent | 5-84 | 0-119 |

### Anchor configuration

In the anchor configuration, all 80 of the input views have been directly compressed with the use of modified 3D‑HEVC encoder version 13.0 (HTM13) [6]. Since normative HTM 13 supports only up to 62 views, the 80 input views were divided into two sets, 40 views each. Each set of 40 views was encoded separately and decoded for the display.

Four rate-points R1 to R4 (Table 2) were specified both for anchors and the proponents. In each case, total bitrate necessary for transmission of all of the views have been specified.

Table 2. Specification of rate-points for SMV case.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Seq. Name | Maximum bitrate for all data for recreation of required  number of output views [kbps]  (QP values used for anchor bitstreams) | | | |
| Rate-point 1 | Rate-point 2 | Rate-point 3 | Rate-point 4 |
| Big Buck Bunny flowers | 5513.0 (35) | 3298.5 (40) | 1905.0 (45) | 1156.3 (50) |
| Big Buck Bunny butterfly | 1665.9 (37) | 1245.3 (40) | 862.9 (44) | 563.9 (50) |

### Preparation of the material for subjective evaluation

The materials to be evaluated were prepared according to the methodology described in the CfE [6] as follows. Since the 80 view 3D display is hard to attain, the evaluation was done in a simulated manner by sweeping through all of the views from right to left and back again. The starting position of the sweeps was selected randomly by the test chair as specified in Table 3. Sweeps were constructed at a speed of one frame per view (Fig. 2).

Table 3. Starting positions of the sweeps, selected randomly by the test chair.

|  |  |  |
| --- | --- | --- |
| No. | Seq. Name | Starting position of the sweeps |
| 3 | Big Buck Bunny flowers | 55 |
| 4 | Big Buck Bunny butterfly | 45 |

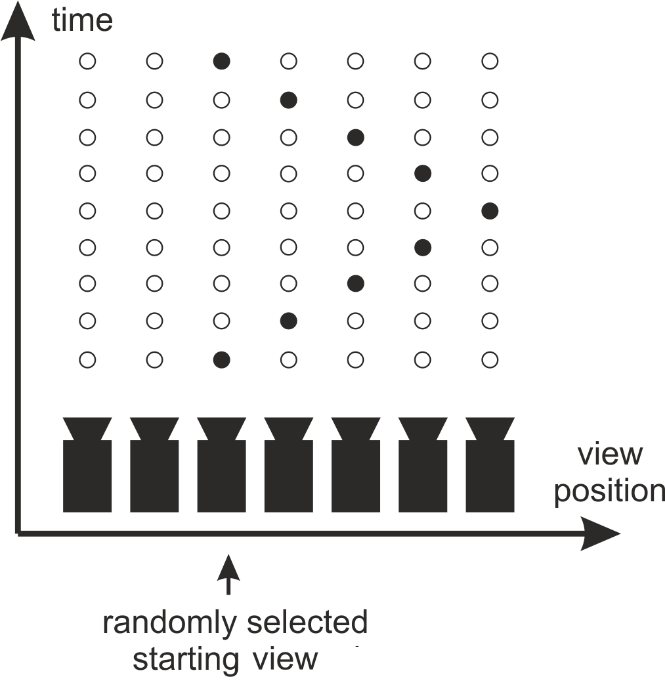


Fig. 2. Super-Multiview (SMV) display application scenario.

In other words, the quality was measured by creating two artificial sequences (for the left and the right eye) that were displayed to a user who started from a randomly selected camera #M and moved to the right until camera #80 is reached and moves back to camera #1 and forth to camera #80. Because the sequences have frame rate of 25 frames per second, the speed at which the user sweeps through the cameras is 25 cameras per second, i.e. 5/7ᵒ/s.

This evaluation procedure has been performed both for the anchor and the response.

## Received responses

There was only one response received [9][10] in SMV category. It conforms all of the requirements of the CfE, but includes only results for two sequences: Big Buck Bunny Flowers and Big Buck Bunny Butterfly. The idea of the response consists in coding of a limited set of views (e.g. 28 out of 80 views for Flowers and 13 out of 80 views for Butterfly) with their corresponding depth maps estimated at the encoder (from the full set of views, e.g. 80). For that, a new algorithm called fDEVS (fast Depth Estimation and View Synthesis) was used. The idea of the proposed pipeline has been illustrated in Figure 3. The estimated depth maps by Fast DERS are encoded (along with the views) with unmodified HTM 13.0. At the decoder, the remaining (not coded) views are synthesized by VSRS with the use of depth-image based rendering (DIBR) from the dectded texture views and the associated depth views.

For Butterfly sequence, the left most view (view5) or the right most view(view84) was synthesized from the decoded view6 or view83, since the depth estimation software (DERS) generally requires the left view and right view to estimate a center depth view. However, the quality of such synthesized views from only one reference view is generally lower than the other views synthesized from both decoded left view and right view. For the Flowers sequence, the left most depth map (view5) and the right most depth map (view84) were synthesized from one depth map of view6 or view83 by the view synthesis software (VSRS) at the encoder side. The quality of such depth maps synthesized from the only one reference view is almost same as the depth maps estimated from both left view and right view, since the depth maps generally have no texture. Because of that, the inpainting tool of fDEVS works well.

For example in the case of Butterfly sequence, 13 texture views and the 13 depth views of Butterfly sequence are encoded by utilizing the correlation between texture and depth since the number of all layers (26) is small. In the case of Flowers sequence, 28 texture views and 28 depth views are encoded separately in order to accelerate the encoding speed.

During the texture view encoding, depth estimation and then depth encoding can be done in parallel. The biggest issue of SMV scenario is how to accelerate the encoding speed.



Fig. 3. fDEVS (fast Depth Estimation and View Synthesis) for Flowers sequence. For Butterfly sequence the scheme is little modified: (1) VSRS is remove from in-between of fast DERS and HTM 13 Enc, (2) number of 28 views and depths is changed to 13, and (3) the two HTM encoders are combined into one, and 4) the HTM decoders are combined into one.

## Evaluation

### Objective evaluation

In Fig.4(a) and (c), each PSNR of the decoded views and the synthesized views are depicted. In Fig.4(b) and (d), average PSNR of all decoded views and synthesized views have been measured. Total bitrate of all data, including the depth maps and zipped supplementary information of the camera parameters, is used.

1. PSNRs of decoded view and synthesized view for Flowers



1. Average PSNR of all 80 views vs total bit rare including depth maps and cam param bits for Flowers

Fig. 4. Results of objective evaluation of the technologies in Super-Multiview (SMV) application scenario.

1. PSNRs of decoded view and synthesized view for Butterfly



1. Average PSNR of all 80 views vs total bit rare including depth maps and cam param bits for Butterfly

Fig. 4 (continued) Results of objective evaluation of the technologies in Super-Multiview (SMV) application scenario.

Table 4 shows Bjøntegaard deltas (BD-RATE) [11] of the response versus anchors (Average PSNR of 80 views / All data bitrate). There were calculated basing on:

1. Anchors sequences:

* total number of bits of the all 80 encoded views,
* average PSNR of all of the 80 decoded views.

1. fDEVS sequences:

* total number of bits of the limited set of encoded views and depth maps,
* average PSNR of the all 80 views reconstructed at the decoder (some of which were directly decoded and some were synthesized from decoded views and depths).

Table 4. Summary of the objective results for Super-Multiview (SMV) application scenario. Bjøntegaard deltas (BD-RATE) [11] of the response versus anchors (Average PSNR of 80 views / All data bitrate). Negative values depict gains of the response.

|  |  |
| --- | --- |
| **Sequence** | **NICT** |
| Big Buck Bunny flowers | -3,2% |
| Big Buck Bunny butterfly | -17.5% |
| **Average (all)** | **-10.4%** |

### Subjective evaluation

Both the response and the anchor for Super-Multiview scenario were evaluated through formal subjective testing. The viewing was performed on June 1st 2016, starting from 8:30am.

The created sweeps were evaluated subjectively with ACR-HR method (Absolute Category Rating with Hidden References) [12] with the use of 11-point MOS (Mean Opinion Score) scale from 0 (bad) to 10 (excellent).

The viewers, 12 in total, were asked to evaluate the quality of the shown video, related to their own developed absolute quality scale. Each session was preceded with a training session in order to stabilize and calibrate the scores given by the viewers.

The presented data consisted in:

* the references (uncompressed original input data- anchor R0),
* the anchors (input data compressed with HTM 13) at four rate-points (anchor R1-R4),
* the response, each at four rate-points (response R1-R4).

Additionally, randomly selected test points have been repeated in order to allow consistency check. The presentation order was randomized.

The vote scores were analyzed statistically which yielded the mean results and 95% confidence intervals.

Fig.5 shows 11 levels of MOS scores for the proponent(blue) and anchor(orange) at four rate points (R1–R4).





Fig. 5. Results of subjective evaluation of Super-Multiview (SMV) application scenario.

For some of the rate-points there were observed statistically significant differences from the anchor e.g. R1, R2 and R4 of BBB Butterfly and R4 of BBB Flowers. For others, although, there is a tendency that no statistically significant conclusion can be made. It should be noted however that the proponent’s bit rates for BBB Butterfly are significantly lower than the anchor at all rate points. Figure 6 interpolates the proponent’s and the anchor’s MOS scores according to the bit rate.

# Application Scenario #2: Free Navigation (FN)

## Overview

The source material in FN applications is a sparse number of views (e.g., 6 - 10) with arbitrary positioning and wide baseline distance between each view. The input views, along with all supplemental data such as depth are transmitted. The output will render arbitrary view positions in 3D space.

The reference framework is comprised of sparse views with large baselines that are not very highly correlated (relative to the ultra-dense views required for Super Multiview displays). The key challenge in the Free Navigation scenario resides in keeping the quality of the view rendering from compressed representation of arbitrarily positioned views notably high. Existing depth-based 3D formats are not able to satisfy such flexible rendering needs with high quality and robustness.

## Objective

The main objective in this application scenario is to substantially improve rendering quality at arbitrary virtual view positions in 3D space from compressed representation of views. It is expected that this may be achieved through an alternative representation format (different from simulcast HEVC and 3D-HEVC), in which case compression efficiency must also be considered. While the emphasis is on the rendering and view synthesis quality, it should be clarified that there is no intention to standardize post-processing tools subsequent to the decoder in the processing chain. However, a more appropriate representation/coding model may be required.

## Methodology

### Test Material

The data sets specified in the respective sections of the Call for Evidence (CfE) [6] were used for the evaluation. The source is a sparse set of 7 views with arbitrary positioning and wide baseline distance between each view. The input views, along with all supplemental data, such as depth, are transmitted. At the output of considered system arbitrary view positions in 3D space can be rendered (Fig. 6).

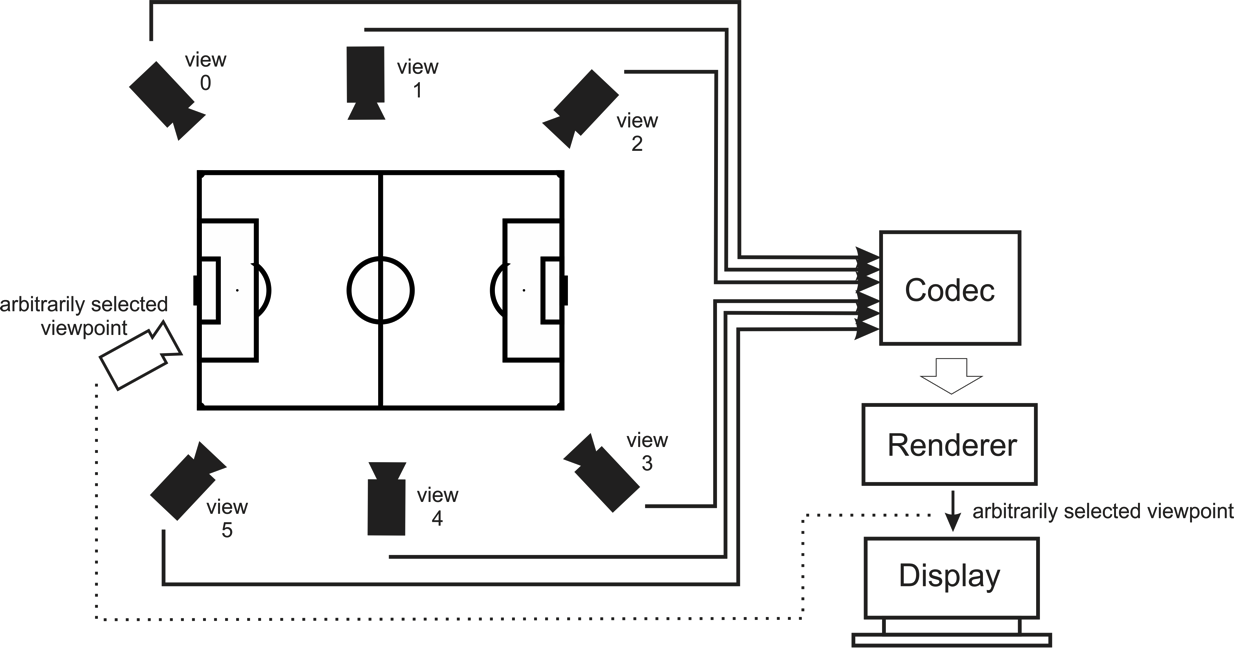


Fig. 6. Free Navigation (FN) application scenario.

Sequences that has been used for evaluation are summarized in Table 5. Seven views (as specified in CfE) of each sequence were coded and used for evaluation. Their positions are also shown in Table 5.

Sequences that are shorter than 10 seconds have been extended twice in a ping-pong loop to last exactly 10 seconds. This has been done immediately after decoding (e.g. before creation of the sweeps).

Table 5. Summary of the sequences used.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. | Source | Seq. Name | Number of Views | Resolution (pel) | Frame rate (fps) | Length | Camera Arrangement | Depth Data Available | Views positions to be transmitted | Frame range to be transmitted |
| 1 | UHasselt | Soccer-Linear 2 | 8 | 1392x1136 | 60 | 10 sec 600 frames | 1D parallel | Yes | 1-7 | 0-599 |
| 2 | UHasselt | Soccer-Arc 1 | 7 | 1920x1080 | 25 | 22 sec 550 frames | 120 deg. Corner, arc | Yes | 1-7 | 0-249 |
| 3 | Poznan University of Technology | Poznan Blocks | 10 | 1920x1080 | 25 | 40 sec 1000 frames | 100 deg. arc around the scene | Yes | 2-8 | 0-249 |
| 4 | Holografika | Big Buck Bunny  Flowers  noBlur | 91 | 1920x1080 | 24 | 5 sec  121 frames | 45 deg, arc | Yes, ground truth depth | 6,19,32,45,58,71,84 | 0-120 |

### Output virtual views and view synthesis

The views, described in Table 6, coded along with the depth data, were reconstructed and used for view synthesis of intermediate virtual views, placed in–between of the input views.

The numbers of required virtual views rendered between each pair of transmitted views are provided in Table 6. Exact camera parameters for each individual virtual viewpoint were taken from the attachment of the CfE.

Table 6. Number of virtual views between each pair of cameras.

|  |  |
| --- | --- |
| Sequence | Number of required virtual views between each pair of cameras |
| Soccer Linear 2 | 14 |
| Soccer Arc 1 | 22 |
| Poznan Blocks | 12 |
| Big Buck Bunny Flowers | 12 |

### Anchor and submission configuration

In the anchor configuration all of the input views along with the corresponding depth data have been compressed using modified 3D-HEVC encoder (based on HTM version 13.0). After decoding, the requested view positions were synthesized using VSRS version 4.1 with enabled depth-based depth blending technique. The details about the anchor configuration are in the CfE.

In anchor configuration seven input views together with their corresponding depth data were encoded at 4 rate-points, as specified in Table 7.

Table 7. Specification of rate-point for Free Navigation (FN) scenario.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Seq. Name | Maximum bitrate for all data required for recreation of required number of output views [kbps] (QP texture/ QP depth) | | | |
| Rate-point 1 | Rate-point 2 | Rate-point 3 | Rate-point 4 |
| 1 | Soccer Linear 2 | 3952.8 (30/39) | 1287.9 (37/43) | 506.0 (44/47) | 362.8 (47/50) |
| 2 | Soccer Arc | 5462.9 (30/39) | 2284.2 (37/43) | 899.1 (44/47) | 592.0 (47/50) |
| 3 | Poznan Blocks | 5927.8 (30/39) | 3187.95 (35/42) | 1559.6 (40/45) | 823.4 (45/48) |
| 4 | Big Buck Bunny Flowers | 1769.7 (37/43) | 1267.8 (40/45) | 816.9 (44/47) | 561.2 (47/50) |

### Preparation of the material for subjective evaluation

The materials for the subjective evaluation were prepared according to the methodology described is the CfE [6] as follows. Video clips of the rendered views were combined to create sweeps through all of the rendered and reconstructed views. The starting positions of the sweeps were selected randomly by the test chair (Table 8).

Table 8. Starting positions of the sweeps selected randomly by the test chair.

|  |  |  |
| --- | --- | --- |
| No. | Seq. Name | Starting position of the sweeps |
| 1 | Soccer Linear 2 | 60 |
| 2 | Soccer Arc | 110 |
| 3 | Poznan Blocks | 20 |
| 4 | Big Buck Bunny Flowers | 35 |

The sweeps (Fig. 7) were constructed at a speed of one frame per view. This has been performed for the anchor and the responses.

Virtual views

virtual views

time

coded  
 views

Virtual path (sweep) with randomly selected starting position

Fig. 7. Free Navigation (FN) evaluation procedure.

## Received responses

There were three responses received which were evaluated in the context of the CfE:

* Poznan University of Technology [13][14].
* Zhejiang University [15][16].
* Hasselt University [17].

### Poznan University of Technology

Response from Poznan University of Technology is based on 3D-HEVC technology (modified HTM version 13.0). However, it is not compatible with 3D-HEVC standard, e.g. proposed bitstreams cannot be decoded with the use of 3D-HEVC decoder. Most of the high level syntax does not changes, but also there are some adjustments at low-level. Several syntax elements and tools have been modified:

* Transmission of camera parameters in VPS, including extrinsic and intrinsic parameters, like rotation matrices, translation and distortion parameters.
* Modification of Disparity Compensated Prediction (DCP). In our codec instead of disparity (along restricted horizontal direction) we use depth-based compensation.
* Modification of Neighboring Block Disparity Vector (NBDV). Instead of disparity restricted to horizontal direction, we use a vector.
* Modification of Depth-oriented NBDV (DoNBDV). In 3D-HEVC the disparity for a given block is set to the value that corresponds with the maximum value of four corner depth samples value of virtual depth map block. In the proposed method, the disparity is calculated based on half of the maximum depth sample value and the position of selected corner of the block.
* Modification of View Synthesis Prediction (VSP). In 3D-HEVC view synthesis in prediction is restricted to horizontal translation only. In our codec, full DIBR scheme is performed.
* Modification of Inter-view Motion Prediction (IvMP). In 3D-HEVC motion vectors are purely 2D. Because other views lay on the same plane, motion vectors after projection to other view remain the same. In the proposed extension, during the prediction, we accordingly rotate motion vector in 3D space.
* Color correction. As an initial preprocessing step the input views are color corrected.

The configuration is resembling those of anchors in the CfE [6], i.e. Main Profile, GOP size = 8, intra period = 24, hierarchical GOPs on, 4 reference frames, Neighboring Block Disparity Vector on, Depth oriented NBDV on, View Synthesis Prediction on, Inter-view Motion Prediction on, Illumination Compensation on.

One difference worth noticing is that View Synthesis Optimization (VSO) has been switched off.

### Zhejiang University

Response from Zhejiang University is based on 3D-HEVC technology (modified HTM version 13.0). However, it is not compatible with 3D-HEVC standard, e.g. proposed bitstreams cannot be decoded with the use of 3D-HEVC decoder. Several coding tools have been modified and improved:

* Disparity Compensated Prediction: depth-based compensation, 2D disparity.
* Neighboring Block Disparity 2D-Vector.
* 2D disparity based View Synthesis Prediction (VSP).
* Adjust DV borrowed from neighboring block according to position of blocks.
* Adjust Depth borrowed from neighboring block according to position of blocks.
* Projection of motion vector as Inter-view MV predictor.
* Depth RDO for non-linear Camera set.

View Synthesis Optimization (VSO) have been adjusted to non-linear camera arrangement.

### Hasselt University

Response from Hasselt University includes results only for Soccer Linear 2 sequence.

In the proposed approach depth data is not necessary to be transmitted and it is estimated at the decoder. Efficient plane sweeping depth estimation is proposed to estimate depth from the decoded video. Because the depth is not transmitted the bitrates of submitted bitstreams are adequately smaller than those of the anchors and thus smaller than the required rate-points R1‑R4.

For the view synthesis, the response uses proprietary view synthesis software.

### Comparison of the submitted technologies

Two of the responses submitted (3.2.1 and 3.2.2) are the extensions of the 3D-HEVC technology for non-linear camera arrangement. Some of the tools are common in the proposals. There are also some tools that are disjoint and are implemented exclusively in only one of the responses.

Merging of the tools from both responses can possibly result in higher compression performance than any of them independently.

The last response (3.2.3) is not proposing any new coding technology but rather focuses on decoder side depth estimation and rendering.

Summary of proposed tools and technologies is made in Table 9.

Table 9. Comparison of the tools submitted in the responses to the CfE in FN category.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tools** | **Poznan University** | **Zhejiang University** | **Hasselt University** |
| Color correction: preprocessing. | Y | - | - |
| Camera parameter transmission extension:  extrinsic and intrinsic parameters, like rotation matrices, translation and distortion parameters. | Y | - | - |
| Disparity Compensated Prediction:  depth-based compensation，2D disparity. | Y | Y | - |
| Neighboring Block Disparity 2D-Vector | Y | Y | - |
| Modification of Depth-oriented NBDV (DoNBDV):  disparity is half of the maximum depth sample value and the position of selected corner of the block. | Y | - | - |
| 2D disparity based View Synthesis Prediction (VSP). | Y | Y | - |
| Adjustment of DV borrowed from neighboring block:  according to position of blocks. | - | Y | - |
| Adjustment of Depth from neighboring block:  according to position of blocks. | - | Y | - |
| Projection of motion vector as Inter-view MV predictor | Y | Y | - |
| Depth RDO for non-linear camera arrangement  View Synthesis Optimization for non-linear camera arrangement | - | Y | - |
| Proprietary view synthesis software:  e.g. based on multiple views, higher precision, hole filling, color correction | Y | - | Y |
| Depth estimation at the decoder side | - | - | Y |

## Evaluation

### Objective evaluation

Average PSNR values of all coded have been measured and used to calculate Bjøntegaard bitrate deltas [11]. The attained values are presented in Tables 10 and 11. Total bitrate of all data including a depth maps and supplementary information was used.

Table 10. Summary of the objective results for Free Navigation (FN) application scenario. Bjøntegaard deltas (BD-RATE) [11] of the proposal versus anchors (Decoded view PSNR / Decoded video bitrate). Negative values depict gains of the proposal.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Bjøntegaard** **delta, versus anchors (Decoded view PSNR / Decoded video bitrate)** | | |
| **Sequence** | **UHasselt\*** | **Poznan** | **Zheijang** |
| Big Buck Bunny flowers | - | -5.21% | -4.62% |
| PoznanBlocks | - | -16.25% | -9.77% |
| SoccerArc | - | -1.14% | -11.93% |
| SoccerLinear | -31,6% | +2.20% | -0.18% |
| **Average (nonlinear)** | - | **-7.53%** | **-8.77%** |
| **Average (all)** | - | **-5.10%** | **-6.62%** |

\*- this proposal consists in not sending the depth data and thus the bitrates of submitted bitstreams are lower, which corresponds to the bitrate reduction.

Table 11. Objective results for Free Navigation (FN) application scenario of Zhejiang University response. Bjøntegaard deltas (BD-RATE) [11] of the proposal versus anchors (decoded and synthesized views PSNR / Total bitrate).

|  |  |
| --- | --- |
|  | **Bjøntegaard** **delta, versus anchors (Decoded and synthesized view PSNR / total bitrate)** |
| **Sequence** | **Zheijang** |
| Big Buck Bunny flowers | -2.08% |
| PoznanBlocks | -11.84% |
| SoccerArc | -17.60% |
| SoccerLinear | -0.04% |
| **Average (nonlinear)** | **-10.51%** |
| **Average (all)** | **-7.89%** |

We consider Soccer linear 2 independently as it is a linearly arranged sequence, and is hence outside the scope of the current FTV activity, which targets multi-camera setups beyond linear arrangements.

### Subjective evaluation

The submissions and the anchor for Free Navigation scenario were evaluated through formal subjective testing. The viewing was performed on May 31st 2016, starting from 2pm.

The created sweeps were evaluated subjectively with ACR-HR method (Absolute Category Rating with Hidden Reference) [12] with the use of 11-point MOS (Mean Opinion Score) scale from 0 (bad) to 10 (excellent).

The viewers, 30 in total, were asked to evaluate the quality of the shown video, related to their own developed absolute quality scale. Each session was preceded with a training session in order stabilize and calibrate the scores given by the viewers.

The presented data consisted of sweeps prepared from the following data:

1. Anchors

* decoded data at four rate-points (anchor R1-R4),
* uncompressed original input data (anchor R0).

1. Responses

* decoded data at four rate-points (response R1-R4),
* uncompressed original input data, synthesized with the submitted view synthesis technology (response R0 if available, anchor R0 otherwise).

In order not to exceed the human focus time, the whole test was divided into two parts. Each part were repeated 3 times resulting in 6 session it total. Additionally, randomly selected test points have been repeated in order to allow consistency check. The presentation order was randomized per session in order to prevent any contextual affect appear.

The attained scores were analyzed statistically which yielded the mean results and 95% confidence intervals. All of those are shown in Figures 8 and 9.

There were observed statistically significant differences from the anchor in the following cases:

* BBB flowers sequence, ratepoints R1, R2 and R3 - proposal from Poznan University of Technology is better than the anchor,
* Poznan Blocks sequence, ratepoints R1 and R3 - proposal from Poznan University of Technology is better than the anchor, in ratepoint R1 proposal from Zhejiang University practically also is better than the anchor,
* Soccer Linear 2 sequence, ratepoint R3 - proposal from Poznan University of Technology is better than the anchor,
* Soccer Arc sequence, ratepoint R1 - proposal from Zhejiang University practically also is better than the anchor, ratepoints R2 – proposal from both Poznan and Zhejiang better than the anchor.
* In all of the mentioned cases, the gain of the proposals in about 1 MOS point. For some other cases there is tendency that no statistically significant conclusions can be made.





Fig. 8. Results of subjective evaluation of Free Navigation (FN) application scenario. The bars related to four coded rate-point R1-R4, as well as synthesis from the uncompressed original (R0) are marked.





Fig. 9. Results of subjective evaluation of Free Navigation (FN) application scenario. The bars related to four coded rate-point R1-R4, as well as synthesis from the uncompressed original (R0) are marked.

The differences in the achieved results in subjective and objective evaluation come from the fact that in objective evaluation only the decoded views are considered, while in the subjective evaluation also the rendering capability was assessed.

# Concluding remarks

The group has evaluated the responses to the CfE. The obtained results show clear improvement and suggest that there is technology better adapted to the considered application scenarios of SMV and FN than the currently standardized solutions.

In Super-Multiview (SMV) application scenario, one response showed gains for two sequences from the set. The observed overall Bjøntegaard bitrate reduction is about 10% objectively.

In Free Navigation (FN) application scenario the observed overall Bjøntegaard bitrate reduction is on average from about 7% to 10% depending on the response, for the set of sequences with nonlinearly arranged cameras, which constitute the main scope of the FTV group. Also subjective evaluation was performed for the responses and in some of the cases statistically significant improvements were observed.

As a side product of the evaluation of the responses it was observed that there is a room for improvement of the quality of the views rendered from the uncompressed data.

# References

[1] <https://www.youtube.com/watch?v=T0VPjHwrc-o>, "Signal Processing in Free Viewpoint Television", IEEE Signal Processing Society.

[2] Masayuki Tanimoto, “FTV (Free-viewpoint Television)”, APSIPA Transactions on Signal and Information Processing, Vol. 1, Issue 1, e4 (14 pages) (August 2012). doi: 10.1017/ATSIP.2012.5.

[3] W. Matusik, H. Pfister, “3D TV: A Scalable System for Real-Time Acquisition, Transmission and Autostereoscopic Display of Dynamic Scenes”, ACM SIGGRAPH, ISSN: 0730-0301, Vol. 23, Issue 3, pp. 814-824, August 2004.

[4] “FTV Seminar Report,” ISO/IEC JTC1/SC29/WG11 MPEG2014/N14552, Sapporo, Japan, July 2014.

[5] “Use Cases and Requirements on Free-viewpoint Television (FTV),” ISO/IEC JTC1/SC29/WG11 MPEG2014/N14178, San Jose, US, January 2014.

[6] “Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation”, ISO/IEC JTC1/SC29/WG11 MPEG2015/N15348, June 2015, Poland, Warsaw.

[7] Vittorio Baroncini, Masayuki Tanimoto, Olgierd Stankiewicz, "Results of the Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation" ISO/IEC JTC1/SC29/WG11 MPEG2016/N16128, San Diego, US, February 2016.

[8] Vittorio Baroncini, Masayuki Tanimoto, Olgierd Stankiewicz, "Summary of the results of the Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation" ISO/IEC JTC1/SC29/WG11 MPEG2016/N16318, Geneve, CH, June 2016.

[9] Takanori Senoh, Akio Ishikawa, Makoto Okui, Kenji Yamamoto, Naomi Inoue, “FTV AHG: CfE Response fDEVS” , ISO/IEC JTC1/SC29/WG11, MPEG2016/ m37953, San Diego, US, February 2016.

[10] Takanori Senoh, Kenji Yamamoto, Nobuji Tetsutani, Hiroshi Yasuda, "FTV AHG: Flowers results with CfE Response fDEVS".

[11] G. Bjøntegaard, “Calculation of Average PSNR Differences between RD-curves”, ITU-T SG16, Doc. VCEG-M33, April 2001.

[12] [ITU-T Rec. P.910 : Subjective video quality assessment methods for multimedia applications](https://www.itu.int/rec/T-REC-P.910/en), 2008.

[13] Marek Domański, Adrian Dziembowski, Adam Grzelka, Łukasz Kowalski, Dawid Mieloch, Jarosław Samelak, Olgierd Stankiewicz , Jakub Stankowski , Krzysztof Wegner, „[FTV AHG] Technical Description of Poznan University of Technology proposal for Call for Evidence on Free-Viewpoint Television”, ISO/IEC JTC1/SC29/WG11, MPEG2016/m37893, San Diego, US, February 2016.

[14] Marek Domański, Adrian Dziembowski, Adam Grzelka, Łukasz Kowalski, Dawid Mieloch, Jarosław Samelak, Olgierd Stankiewicz, Jakub Stankowski, Krzysztof Wegner, "[FTV AHG] Extended results of Poznan University of Technology proposal for Call for Evidence on Free-Viewpoint Television", ISO/IEC JTC1/SC29/WG11 MPEG2016/ m38246, Geneve, CH, May 2016.

[15] Lu Yu, Qing Wang, Ang Lu, Yule Sun, “Response to Call for Evidence on Free-Viewpoint Television: Zhejiang University”, ISO/IEC JTC1/SC29/WG11, MPEG2016/m37608, San Diego, US, February 2016.

[16] Ang Lu, Yule Sun, Lu Yu, "Extended results of ZJU’s CfE on Free-Viewpoint Television" ISO/IEC JTC1/SC29/WG11 MPEG2016/M38549, Geneva, Switzerland, May 2016.

[17] Lode Jorissen, Patrik Goorts, Gauthier Lafruit, “[FTV AhG] Soccer Light Field Interpolation Applied on Compressed Data”, ISO/IEC JTC1/SC29/WG11, MPEG2016/m37674, San Diego, US, February 2016.