

# New visual coding exploration in MPEG: Super-MultiView and Free Navigation in Free viewpoint TV

Gauthier Lafruit, Université Libre de Bruxelles (Belgium); Marek Domański, Krzysztof Wegner and Tomasz Grajek, Poznań University of Technology (Poland); Takanori Senoh, National Institute of Information and Communications Technology (Japan); Joël Jung, Orange Labs (France); Péter Tamás Kovács, Holografika (Hungary); Patrik Goorts and Lode Jorissen, Hasselt University; Adrian Munteanu and Beerend Ceulemans, Vrije Universiteit Brussel (Belgium); Pablo Carballeira and Sergio García, Universidad Politécnica de Madrid (Spain); and Masayuki Tanimoto, Nagoya Industrial Science Research Institute (Japan)

## Abstract

ISO/IEC MPEG and ITU-T VCEG have recently jointly issued a new multiview video compression standard, called 3D-HEVC, which reaches unprecedented compression performances for linear, dense camera arrangements. In view of supporting future high-quality, auto-stereoscopic 3D displays and Free Navigation virtual/augmented reality applications with sparse, arbitrarily arranged camera setups, innovative depth estimation and virtual view synthesis techniques with global optimizations over all camera views should be developed. Preliminary studies in response to the MPEG-FTV (Free viewpoint TV) Call for Evidence suggest these targets are within reach, with at least 6% bitrate gains over 3D-HEVC technology.

## Introduction

Since 25 years, MPEG has steadily been involved in the development of video coding technologies. Today, the most advanced single camera view coding standard, called HEVC (High Efficiency Video Coding) offers a data rate reduction of two orders of magnitude compared to uncompressed video. This provides means to transmit Full-HD TV (High Definition) and soon UHD TV (Ultra High Definition) over communication channels with bitrates of around 15 Mbit/s, ensuring wide acceptance by the general public in the near future.

Over the last decade, ISO/IEC MPEG and ITU-T VCEG have also jointly developed multiview video coding standards (MV-AVC, MV-HEVC) focusing on the compression of multiple camera feeds “as is”, i.e. without means to facilitate the generation of additional views that are not transmitted to the receiver. Depth-based 3D formats – and in particular 3D-HEVC, standardized in February 2015 - have been developed to address this shortcoming: with the use of Depth Image Based Rendering (DIBR) techniques, the generation of additional views from a small number of transmitted views was enabled, supporting glasses-free/auto-stereoscopic 3D display applications with dozens of output views from only a handful of input camera feeds. For example, 3 input  $\rightarrow$  9 output and 5 input  $\rightarrow$  28 output Horizontal Parallax Only (HPO) glasses-free 3D displays have reached the prosumer market, while Super-Multi-View (SMV) light field displays with hundreds of ultra-dense output views and smooth motion parallax are prototyped in R&D labs, e.g. Figure 1.

Unfortunately, aiming at very high quality viewing over a large field of view, one would have to foresee a high number of densely arranged input cameras, reaching 3D-HEVC bitrates in the order of hundreds of Mbit/s for SMV at home cinema quality levels, which might eventually hamper consumer market penetration.



Figure 1. Light Field display (Courtesy of Holografika)

Similarly, in a Virtual Reality (VR) context using Head Mounted Displays (HMD), literally surrounding the scene to visualize with an ultra-dense arrangement of several hundreds of cameras would indeed offer correct motion parallax and Free Navigation (FN) functionalities around the scene (cf. Figure 2), similar to the Matrix bullet effect. Additionally, zoom-in/out functionalities (cf. arrow 4 in Figure 3) would extend the walk-around feeling to a truly immersive “fly through the scene” VR experience on authentic looking content.

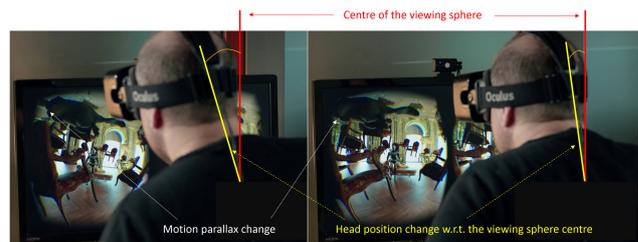


Figure 2. Motion parallax in Virtual Reality (Courtesy of Nozon)

However, to fully enable take-up of such VR technology in each living-room, drastic cost reductions in multi-camera content acquisition and transmission should be reached, which inevitably calls for a reduction in the number of acquisition cameras and the development of high-performance DIBR virtual view synthesis techniques with sparse camera arrangements.

3D-HEVC being primarily developed for consumer autostereoscopic 3D displays in linear camera arrangements with small inter-camera distance (small/narrow baseline), new compression and view synthesis challenges have to be explored for

the aforementioned Super-MultiView (SMV) and Virtual Reality Free Navigation (VR-FN) application scenarios with moderately dense or sparse, arbitrarily arranged multi-camera setups. MPEG therefore recently issued a Call for Evidence (CfE), calling companies and organizations to demonstrate technology that they believe perform better than 3D-HEVC and accompanying pre/post-processing. The present paper briefly summarizes the process, the challenges and the expected outcomes for this future standard that – in the absence of an agreed naming in the standardization committee at the time of writing – will be referred in the present paper by 3D-HEVC++ (a naming convention borrowed from C++ that reaches one step further over the well-established C programming language).

### Free Navigation technology by 2020

With respect to the MPEG CfE, the deadline of submission has been settled to 17 February 2016, with an evaluation of the proponents’ responses by the MPEG Free viewpoint TV (MPEG-FTV) Ad-hoc Group during the 114<sup>th</sup> MPEG meeting in San Diego, 20-26 February 2016.

If any of the proposed technologies significantly outperforms currently available MPEG technology, MPEG plans to issue a Call for Proposals (CfP), subsequent to this CfE, to develop standards that offer increased compression performance and viewing experiences beyond 3D-HEVC in SMV and FN application scenarios.

During this development, it is expected that the Olympic Games of Rio de Janeiro in 2016 will bootstrap Multiview coding technologies with discrete multi-viewpoint rendering experiences in many sports events. However, the current view synthesis techniques proposed in MV-HEVC and 3D-HEVC are only competitive in narrow baseline camera setups. It is therefore expected that Free viewpoint TV, allowing the user to navigate freely in the space surrounded by a sparse set of fixed cameras, will need an additional 3-4 years cycle of development before reaching the necessary quality standards at the Olympic Games of Tokyo in 2020. This timeline is well synchronized with the MPEG-FTV CfE and expected CfP schedules.

Moreover, [1] forecasts that VR with multi-camera captured content will represent a \$30 billion market by 2020, with 20% VR films and 45% covering VR games. Already 170 million VR gamers are expected worldwide by 2018 with an annual VR gaming revenue of \$8.6 billion, equally divided over hardware and software. The study also pinpoints the need to develop new image capture and processing technologies (aka Computational Imaging) to overcome the limitation of the user looking around (360 degrees video) from the perspective of the camera’s position only, without any capability to navigate freely within the scene. The technology to allow such Free Navigation (FN) is believed to be based on light field capture [2], which is in line with the multi-camera approach proposed in MPEG-FTV (MPEG Free viewpoint TV), further studied in a newly established Light Field Ad-hoc Group in MPEG [3], as well as in other standardization committees like JPEG-PLENO [4].

### 3D-HEVC extensions for SMV and FN

Figure 3 shows a generic multi-camera setup for real-life application scenarios, with extensions on the current 3D-HEVC codec architecture to support the newly proposed non-linear SMV and/or sparse FN camera arrangements. This should lead to an agile Multiview+Depth transmission scheme, referred to as 3D-HEVC++. The solid line cameras correspond to physical cameras that are setup around the scene, typically in a non-linear

arrangement. The eye icons correspond to user requested virtual viewpoints for which no physical camera views exist. Depth range cameras might also be present to deliver meta-data to the DIBR processing pipeline for virtual view generation, performed in the VSRS (View Synthesis Reference Software) module [5]. The depth meta-data might also be obtained directly from the color cameras through DERS (Depth Estimation Reference Software) [6]. DERS and VSRS are non-normative modules, but nevertheless play an important role in the codec quality-bitrate performance figures, calling for their in-depth study and improvement in the development of the future 3D-HEVC++ standard.

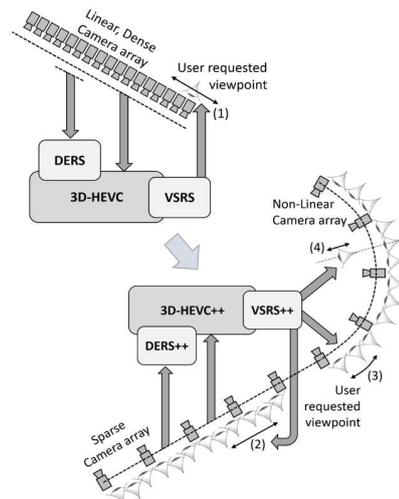


Figure 3. Multiview plus depth video pipeline for 3D-HEVC (top-left) and 3D-HEVC++ (bottom-right) showing the input cameras and user requested views (eyes) that are synthesized along linear (1, 2) and non-linear/curved pathways (3), as well as zoom-in/out functionalities (4) to obtain viewpoints within the enclosed camera volume.

Indeed, the (optional) depth maps are compressed together with the color images, and view synthesis is also used during compression in order to provide a prediction to a physical camera view from its direct neighbors for transmitting a low-entropy difference image to the receiver. This View Synthesis Prediction (VSP) is a codec-in-the-loop method, hence will not impact the decoded view quality in case of an imperfect view synthesis (though it will then increase the bitrate). However, an additional view synthesis (VSRS) step will be applied from the decoded views to generate additional virtual viewpoints that are not transmitted to the receiver. Since this view synthesis works in an open-loop mode, any artefact in the generated views will have a dramatic impact on the perceived output quality. This is an important reason to explore new view synthesis techniques that can work properly in large baseline conditions.

Finally, it is worth noticing that in an SMV display, all the physical and virtual camera viewpoints have to be rendered simultaneously. In a VR-FN application scenario, however, only two adjacent viewpoints (physically existing and/or virtual viewpoints) have to be rendered at any given moment in time in the stereo HMD, based on the user’s current position. Since VR does not tolerate high response latencies, the complexity of the employed techniques should remain acceptable.

### SMV and FN test sequences

The MPEG-FTV group recommends specific SMV and FN test sequences in the MPEG-FTV CfE, in order to conduct comparative studies between the submitted technologies [7]. The SMV sequences contain 80 narrow-baseline views, while the FN sequences contain only 7 views, each view being complemented by a depth map that has been estimated offline, either by DERS, or by a proponent's in-house technique.

The Big Buck Bunny SMV sequences are generated from 3D graphics files donated by the Blender Foundation. Eighty adjacent viewing directions were synthetically rendered by Holografika to obtain the Big Buck Bunny color and depth map videos used in the CfE evaluation. Seven of these views are also used as sparse FN sequences (Flowers, Butterfly). The Big Buck Bunny Flowers and Butterfly depth maps do not contain any artefacts, since they are synthetically rendered from a 3D model by conversion from the z-buffer during rendering. However, the depth maps of all other sequences (Champagne Tower, Pantomime, Soccer-Arc1, Soccer-Linear2 and Poznan Blocks) have been estimated algorithmically (DERS or proprietary software) and show some artefacts, possibly impeding the subsequent view synthesis quality. For instance, since DERS uses a Graph Cut stereo matching technique [8] applied pairwise on adjacent physical camera views, some spatial inconsistencies might appear during view synthesis (VSRS) of virtual views; these are even more apparent in large baseline setups, as will be discussed later in Figures 7 and 8.

### 3D-HEVC in non-linear, large-baseline conditions

The 3D-HEVC technology standardized in February 2015 has originally been developed and tested for linear, narrow baseline camera arrangements. In contrast, convergent cameras in the typical 3D-HEVC++ coding pipeline of Figure 3 will create both positive and negative disparities, cf. Figure 4, requiring to bring minor format and syntax changes into the codec specifications.

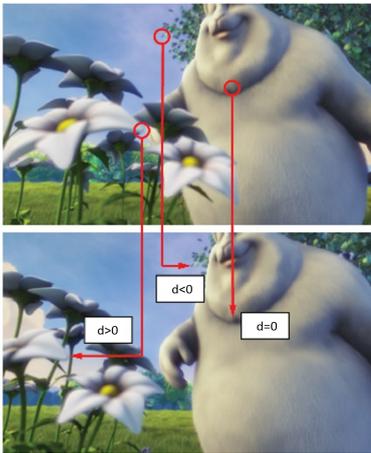


Figure 4. Positive and negative disparities ( $d$ ) in convergent camera setup

Also more essential codec modifications will be required in the development of 3D-HEVC++. For instance, an increase of a distance between cameras results in the reduction of inter-view correlation. This yields deterioration of the 3D-HEVC compression performance that converges to a drastically lower HEVC simulcast compression efficiency. Moreover, for the viewpoint located outside the connecting line between camera views, the interview-prediction

model should be more complex than that based on simple compensation of the horizontal disparity, as it is currently implemented in 3D-HEVC reference software [9].

It is hence expected that new coding developments will be needed, including even non-normative DERS and VSRS developments, which will eventually ripple into the normative 3D-HEVC++ codec specifications. The boundaries between normative and non-normative extensions of 3D-HEVC are consequently gradually blurring away, considerably adding complexity to the 3D-HEVC++ developments.

For instance, View Synthesis Prediction (VSP) is the process of predicting a physical camera view from its two adjacent neighbors, and transmitting the entropy-coded difference image. [10] reports an average bitrate gain of 6.25% over 3D-HEVC by back-and-forth projection between the respective 2D views and 3D space, in non-linear, large baseline camera arrangements. More generally, the implementation of the 3D-HEVC++ should hence exploit the modified disparity vector derivation in such tools as the View Synthesis Prediction (VSP), Disparity Compensated Prediction (DCP), Neighboring Block Disparity Vector (NBDV), Depth oriented NBDV (DoNBDV), Interview Motion Prediction (IvMP) and Illumination Compensation (IC).

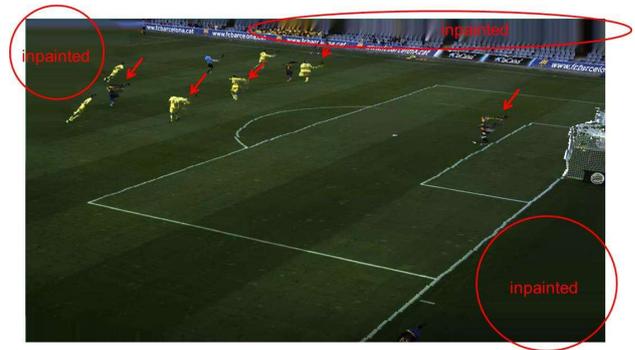


Figure 5. Homography inpainting for VSP in soccer sequence

In extended scenes like soccer fields, even more elaborated techniques are required, cf. Figure 5. For instance, [11] proposes a homography reprojection and inpainting technique for VSP, correcting mainly the outside borders of the camera views. This extension towards novel 3D-HEVC++ technology for arbitrary camera positions is still under development.

### DERS and VSRS in non-linear, large-baseline conditions

As already mentioned, though DERS and VSRS are non-normative in the codec processing pipeline of Figure 3, their performance has an important impact on the quality-bitrate performance figures (e.g. the VSP tool discussed in previous section) and hence also on future developments and updates of the 3D-HEVC codec towards 3D-HEVC++. We therefore give an overview of some improvements that have been studied over the past year in the MPEG-FTV group in supporting the new SMV and FN application scenarios for 3D-HEVC++.



Figure 8. VSRS (top) vs. Epipolar Plane Imaging (bottom) view synthesis

### Large Baseline View Synthesis

In order to serve autostereoscopic- and light field displays with real-life video, advanced view synthesis technologies are needed as it is often impractical to record the high number of camera views that are requested by such displays. The main idea for increasing the compression performance consists in not transmitting some physical camera views at all (in contrast to VSP which transmits a difference image) and generate the missing views with VSRS. For example, in Figure 6, skipping some views during transmission will effectively decrease the bitrate with a factor of 2 in the successive skipping tests (horizontal arrows), but unfortunately the corresponding VSRS generated views also cause a large PSNR drop (4-6 dB in the example of the Champagne Tower test sequence) yielding suboptimal PSNR-bitrate curves.

Figure 7 shows a more detailed view of the PSNR quality degradation when performing an open-loop VSRS view synthesis to recover all output views from a dyadic decreasing number of transmitted views. From the sixty middle views under test, an increasing number of views are not transmitted to the receiver (skip1, skip3, skip5, etc) but rather generated through VSRS. One clearly observes the huge quality degradation of up to a dozen of dB in terms of PSNR for large baselines (high skip numbers). Figure 8 shows the typical horizontal stripes artefacts that are caused at increasing baselines in the VSRS reference software.

Clearly, more in-depth studies are required to evaluate the potential of skipping some views in large baseline scenarios, not only in SMV applications, but foremost in FN applications where VSRS will remain an open-loop tool without error correction post-processing capabilities.

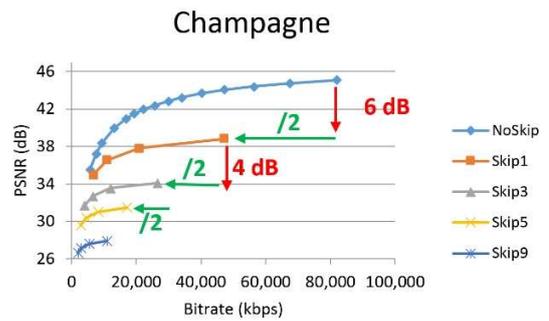


Figure 6. PSNR vs. bitrate results for different coding configurations of the Champagne sequence, by skipping views before the coding (Skip<n>: n consecutive views are skipped and need to be synthesized, between 2 transmitted views)

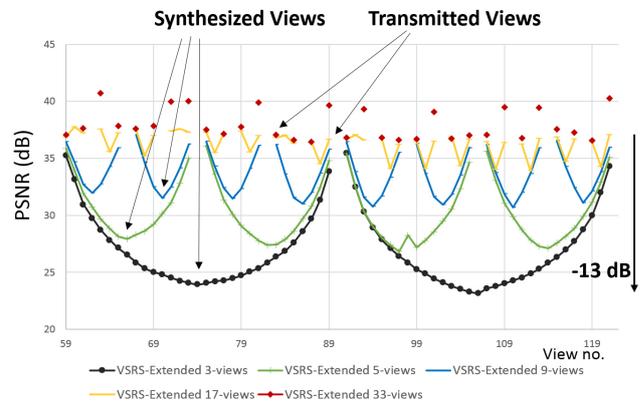


Figure 7. PSNR variation of synthesized views vs. transmitted views



Figure 9. View synthesis with Depth-based view blending (left) vs. VSRS view blending (right)

Recently, some modifications in the VSRS software have been proposed to exclude object depth contributions that are not visible in all camera views, largely improving the view synthesis as shown in Figure 9.

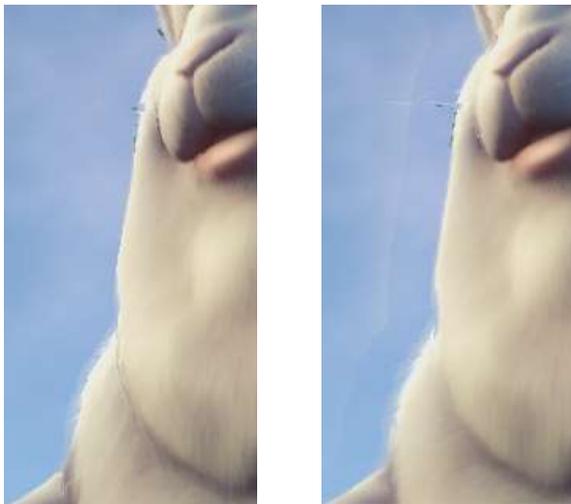


Figure 10. View synthesis without ghosting (left) vs. VSRS (right)



Figure 11. Globally optimized view synthesis (left) vs. VSRS (right)

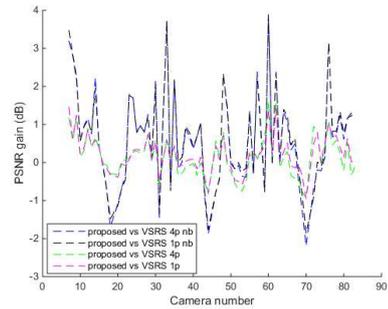


Figure 12. Objective comparison against VSRS on the Big Buck Bunny sequence. Results are reported for single (1p) versus quarter pixel (4p) precision in the warping and turning the view blending option on or off (nb).

Moreover, [12] demonstrates additional improvements to VSRS. Firstly, the algorithm used to perform 3D warping between camera views has been modified in order to avoid ghosting artefacts, cf. Figure 10. Secondly, a new inpainting algorithm is proposed in order to fill disoccluded regions in the image by optimizing a Markov random field using a form of priority-belief propagation [13]. The inpainting algorithm analyzes the depth map in the synthesized view and is designed to reconstruct the disoccluded area using image patches from background regions. Figure 11 clearly shows visual improvements with respect to the current VSRS result. In terms of objective quality expressed by the PSNR, average gains of 0.64dB have been measured for the Big Buck Bunny Flowers sequence. Average PSNR values over time are shown in Figure 12 for each camera in the array.

### Multi-Camera Depth Estimation

Thanks to the techniques described in previous section and the perfect depth map of the Big Buck Bunny Flowers sequence, we have observed that with this test sequence, when skipping a limited number of views (skip1, skip3), the PSNR-bitrate curves remain roughly Pareto optimal with large bitrate gains, as shown in Figure 13.

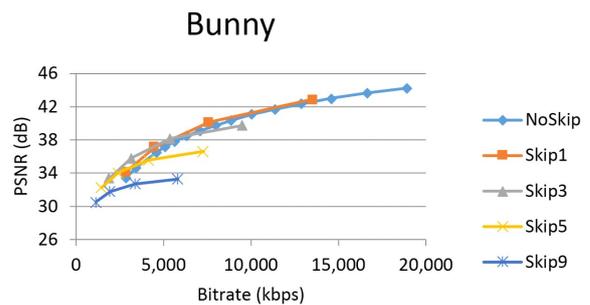


Figure 13. PSNR vs. bitrate results for different coding configurations of the Bunny sequence, by skipping views before coding

For this particular case, the view skipping method as described in the previous section remains interesting, in contrast to the severe PSNR drops observed for large baselines in Figure 7. This is, however, believed to be an exceptional case made possible by the use of perfect depth maps, synthetically calculated for the Big Buck Bunny sequence, hence avoiding further VSRS artefacts induced by depths error.



Figure 14. Depth Estimation (top) and View Synthesis results (bottom) with Segmentation-guided Plane Sweeping (left) and DERS/VSRS (right)

Recent studies show indeed that there is an intricate relationship between the depth and view synthesis distortion in the current DERS and VSRS tools. In particular, [14] provides an exhaustive analysis of the correlation between depth distortion and synthesis distortion at different coding levels, concluding that depth coding distortion reflects well the synthesis distortion at the frame level and MB-row level, while lower correlation values are achieved at the MB level. This analysis also reveals that the distortion on a depth block is aggregated better with a lower-degree norm, Sum of Absolute Error (SAE), than the commonly used Sum of Squared Error (SSE).

In [15], the authors propose a synthesis distortion metric to optimize the coding of depth in coding schemes such as 3D-AVC, 3D-HEVC and 3D-HEVC++. This metric enhances the overall coding efficiency at the cost of a computational complexity overhead introduced by the new metric itself, and the fact that it requires joint processing of depth and texture in a single encoder.

Designing better depth estimation techniques than the current DERS hence provides interesting perspectives to improve view synthesis. In particular, depth estimation based on all available camera views instead of only a subset of them will intuitively be beneficial. In the example of Figure 8, an Epipolar Plane Image (EPI) depth estimation technique [16] using all available camera views, inspired by [17], provides indeed better depth maps and view synthesis results at large baselines, with a 5 dB PSNR gain [16], compensating the typical 4-6 dB losses observed when skipping views in Figure 6. Also [18] reports valuable gains using similar EPI techniques. Finally, [19] has shown large view synthesis subjective quality gains using a segmentation-guided plane-sweep depth estimation method on the Soccer-Arc1 test sequence, cf. Figure 14.

Improving DERS so as to include all available input cameras in the depth estimation, in conjunction with improving VSRS with the techniques described in previous subsection, are clearly interesting directions to be further investigated in the 3D-HEVC++ CfE and subsequent CfP.

## Subjective evaluation of SMV and FN

Though objective quality evaluations based on PSNR give good indications on the most promising candidate compression tools, measuring the Quality of Experience (QoE) plays a crucial role in the determination of the technologies that are adopted in the final standard [20]. For 2D images and video, the well-known ITU-R BT.500-11 recommendation [21] describes the methodology that should be used when performing subjective quality studies involving human participants. In [22], an extension of these guidelines is proposed for the evaluation of 3D content on stereoscopic and multiview autostereoscopic displays.

It is important to notice that SMV and FN content, and their visualization on 3D displays, place new challenges on subjective evaluation of MPEG-FTV coding technology. Some works [23, 24] have helped to provide a parametrization that describes the relations between content, display mode and user experience. Such a parametrization is a very valuable tool to guide the subjective evaluation or even content creation, giving guidelines to configure scene parameters such as depth or density of cameras for an acceptable viewing experience. Particularly, [24] proposes an approach to this parametrization which captures new elements that are relevant in the subjective evaluation of SMV and that do not apply on the evaluation of 2D or fixed-viewpoint stereoscopic video. The main advantage of this novel parametrization is that it is based on the disparity between adjacent views, instead of angle or camera distance, and thus:

- It aggregates the contribution of different parameters that influence the MPEG-FTV subjective experience, better representing the perception of visual comfort.
- It is common to different camera arrangements, such as linear, non-linear convergent or arc.

In particular, such parametrization has been very useful in defining the minimum comfortable camera density in a view path

for the FN scenarios, setting the number of intermediate virtual viewpoint positions between physical cameras [24].

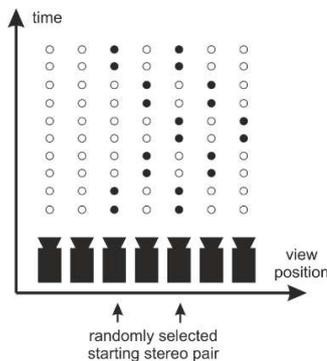


Figure 15. View-sweeping scheme for the stereoscopic evaluation of SMV content in the CfE on FTV.

### CfE stereoscopic viewing

In the CfE process, submissions will be evaluated on a stereoscopic monitor and spatial back-and-forth view sweeps between the left- and right-most views will be generated from the decoded and generated virtual views, cf. Figure 15. Test participants will then provide a Mean Observation Score (MOS) comparing the different technology submissions.

### Light Field SMV viewing

Since subjective viewing on stereoscopic, auto-stereoscopic and light field displays [25] might be very different, rendering quality evaluations should be conducted on a multitude of displays in order to evaluate the best compression technology amongst the CfE proponents.

[26] has shown a linear quality relationship between stereoscopic and auto-stereoscopic displays, but no clear studies are available between the latter and SMV light field displays. Furthermore, to accurately evaluate visual quality in 3D video, it is of paramount importance to avoid any possible visual artifacts introduced by the display's internal light field transmission system, which has to use Gbps communication lines to transmit raw data. To make this possible, Holografika has built a custom light field display of 73 MPixel, with a 2D equivalent resolution of 1280x720 pixels, 24 bit RGB, 70 degrees field of view and an angular resolution of 0.96 degrees, using cluster nodes over a 40GBs Ethernet switch [27]. This system is located at the Electronics and Informatics Department (ETRO) of the Vrije Universiteit Brussel (VUB) in Brussels, Belgium. Raw light field data transport provided by this system offers the possibility to carry out visual tests in MPEG-FTV CfE and subsequent CfP. To this end, the testing environment at VUB-ETRO's 3D Lab has also been equipped with appropriate lighting conditions (non-flickering lights with controllable temperature, specific environmental color), as requested by the ITU-R BT.500-11 methodology for subjective assessment of picture quality [21].

### Conclusion

In order to support Super-MultiView and Free Navigation application scenarios with mostly sparse and/or arbitrarily arranged multi-camera setups, innovative 3D-HEVC extensions should be developed. Preliminary experiments show that the severe quality degradation under large baseline conditions of the MPEG-FTV

VSRS view synthesis can be compensated with global optimization and view synthesis techniques involving all camera views with epipolar plane imaging or plane sweeping techniques. Moreover, better exploiting the non-horizontal-only modified disparity vector derivation in the different coding tools is expected to bring at least 6% bitrate coding gains over 3D-HEVC. Such improvements make applications that generate additional virtual views from a cost-effective multi-camera system viable towards the future.

### References

- [1] Philip Lelyveld, "Virtual Reality Primer with an Emphasis on Camera-Captured VR," Entertainment Technology Center, July 2015, <http://www.etccenter.org/wp-content/uploads/2015/07/ETC-VR-Primer-July-2015o.pdf>
- [2] Mike Seymour, "Light fields – the future of VR-AR-MR," fxguide, 26 May 2015, <https://www.fxguide.com/featured/light-fields-the-future-of-vr-ar-mr/>
- [3] \_, "List of AHGs Established at the 113th Meeting in Geneva," ISO/IEC JTC1/SC29/WG11 MPEG2015/N15622, Geneva, Switzerland, October 2015.
- [4] \_, "JPEG PLENO Abstract and Executive Summary," 20 March 2015, [https://jpeg.org/items/20150320\\_pleno\\_summary.html](https://jpeg.org/items/20150320_pleno_summary.html)
- [5] Krzysztof Wegner, Olgierd Stankiewicz, Masayuki Tanimoto, Marek Domanski, "Enhanced View Synthesis Reference Software (VSRS) for Free-viewpoint Television," ISO/IEC JTC1/SC29/WG11 MPEG2013/M31520, Geneva, Switzerland, October 2013.
- [6] Krzysztof Wegner, Olgierd Stankiewicz, Masayuki Tanimoto, Marek Domanski, "Enhanced Depth Estimation Reference Software (DERS) for Free-viewpoint Television," ISO/IEC JTC1/SC29/WG11 MPEG2013/M31518, Geneva, Switzerland, October 2013.
- [7] \_, "Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation," MPEG 113th meeting, contribution M37296, Geneva, Switzerland, October 2015.
- [8] Yuri Boykov and Vladimir Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), pp. 1124 – 1137, September 2004.
- [9] M. Domański, A. Dziembowski, D. Mieloch, A. Łuczak, O. Stankiewicz, K. Wegner, "A Practical Approach to Acquisition and Processing of Free Viewpoint Video", 31st Picture Coding Symposium PCS 2015, Cairns, Australia, pp. 10-14, 2015.
- [10] Jakub Stankowski, Łukasz Kowalski, Jarosław Samelak, Marek Domański, Tomasz Grajek, Krzysztof Wegner, "3D-HEVC Extension for Circular Camera Arrangements," 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 3DTV- Con 2015, Lisbon, Portugal, 8-10 July 2015.
- [11] T. Senoh, A. Ishikawa, M. Okui, K. Yamamoto, N. Inoue, "FTV AHG: Soccer Arc I Homography Prediction Results", MPEG 113th meeting, contribution M37296, Geneva, Switzerland, October 2015.
- [12] Beerend Ceulemans, et. al., "Efficient MRF-Based disocclusion inpainting in multiview video," submitted to ICME 2016.
- [13] Komodakis, Nikos, and Georgios Tziritas. "Image completion using efficient belief propagation via priority scheduling and dynamic pruning." IEEE Transactions on Image Processing, vol. 16, no. 11 pp. 2649-2661, 2007.

- [14] P. Carballeira, J. Cabrera, F. Jaureguizar, N. Garcia, "Analysis of the depth-shift distortion as an estimator for view synthesis distortion", *Signal Processing: Image Communication*, (accepted on Dec. 2015), <http://dx.doi.org/10.1016/j.image.2015.12.007>
- [15] B. Oh, J. Lee and D. Park, "Depth Map Coding Based on Synthesized View Distortion Function," *IEEE Journal of Selected Topics in Signal Processing*, vol.5, no.7, pp.1344-1352, Nov. 2011.
- [16] Lode Jorissen, Patrik Goorts, Sammy Rogmans, Gauthier Lafruit, Philippe Bekaert, "Multi-Camera Epipolar Plane Image Feature Detection for Robust View Synthesis," *Proceedings of the 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1-4, 2015.
- [17] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, Markus Gross, "Scene Reconstruction from High Spatio-Angular Resolution Light Fields," *ACM Siggraph*, vol. 32, no. 4, 2013.
- [18] Catarina Brites, Joao Ascenso, Fernando Pereira, "Epipolar plane image based rendering for 3D video coding," *IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1-6, October 2015.
- [19] Goorts Patrik, Bekaert Philippe, Lafruit Gauthier, "Real-time, Adaptive Plane Sweeping for Free Viewpoint Navigation in Soccer Scenes," PhD thesis, Hasselt University, 2014.
- [20] Dricot, Jung, Cagnazzo, Pesquet-Popescu, Dufaux, Kovacs, Kiran Adhikarla, "Subjective Evaluation of Super Multi-View Compressed Content on High End Light Field 3D Display", *Signal Processing: Image Communication*, Elsevier, June 2015.
- [21] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," January 2012.
- [22] Lewandowski, Filip, et al., "Methodology for 3D Video Subjective Quality Evaluation," *International Journal of Electronics and Telecommunications*, vol. 59, no. 1, pp. 25-32, 2013.
- [23] P. Carballeira, J. Gutiérrez, F. Morán, J. Cabrera, N. García, "Subjective Evaluation of Super Multiview Video in Consumer 3D Displays", *Seventh International Workshop on Quality of Multimedia Experience, QoMEX 2015, Costa Navarino, Greece*, pp. 1-6, 26-29 May 2015.
- [24] P. Carballeira, J. Gutiérrez, F. Morán, N. García, "New view-sweep parametrization and subjective evaluation of SMV content", *ISO/IEC JTC1/SC29/WG11 MPEG2015/M36448*, Warsaw, Poland, June 2015.
- [25] Kovács, Péter Tamás, et al., "Quality measurements of 3D light-field displays," *Proc. Eighth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*. 2014.
- [26] Krzysztof Wegner, Tomasz Grajek, Marek Domański, "Comparison of 3D video subjective quality evaluated using polarisation and autostereoscopic displays," *Electronics Letters*, Vol. 50, No. 18, pp. 1283-1285, August 2014.
- [27] Kovacs, Peter Tamas, et al., "Analysis and optimization of pixel usage of light-field conversion from multi-camera setups to 3D light-field displays," *IEEE International Conference on Image Processing (ICIP)*, pp. 86-90, October 2014.

## Author Biography

*Gauthier Lafruit is Professor at l'Université Libre de Bruxelles, Brussels, Belgium, in the Laboratory for Image, Signal and Audio processing (LISA). He received his Ph.D. degree in Electrical Engineering from the Vrije Universiteit Brussel, Brussels, Belgium, in 1995. His current research includes Virtual Reality from camera captured content, Light Fields, Computational Imaging and GPU acceleration. He is currently co-chair of the MPEG-FTV group.*

*Marek Domański is a Professor with the Poznań University of Technology, where he leads the Chair (Department) of Multimedia Telecommunications and Microelectronics. He is the author or co-author of six books and over 300 research papers in journals and conference proceedings. His contributions were mostly on image, video and audio compression, image processing, multimedia systems 3-D video and color image technology, digital filters, and multidimensional signal processing.*

*Krzysztof Wegner received the M.Sc. degree from the Poznań University of Technology, Poznań, Poland, in 2008, where he is currently pursuing the Ph.D. degree. He is the co-author of several papers on free view television, depth estimation, and view synthesis. He is involved in ISO standardization activities where he contributes to the development of future 3-D video coding standards.*

*Tomasz Grajek received the M.Sc. and Ph.D. degrees from the Poznań University of Technology, Poznań, Poland, in 2004 and 2010, respectively. He is the author or co-author of several papers on digital video compression, entropy coding, and modeling of advanced video encoders. He has been taking part in several projects for industrial research and development.*

*Takanori Senoh received the Ph.D. degree in Engineering from the University of Tokyo, Japan, in 2007. He is currently with National Institute of Information and Communications Technology, Japan and his current research interests include 3D image processing and electronic holography. He is a member of IEEE, ITE, IIEEJ, and JSAP.*

*Joël Jung received the Ph.D. degree in Electrical Engineering from the University of Nice-Sophia Antipolis, Nice, France, in 2000. He is currently with Orange Labs Paris and B<>Com Institute of Research and Technology, and his current research interests include next generation image and video coding, 3D super multi-view and depth coding. He is an active contributor to the HEVC standard (JCT-VC) and the 3D-HEVC annex (JCT-3V).*

*Péter Tamás Kovács has been working at Holografika since 2006, contributing to the development of the real 3D light-field display product line HoloVizio and related technologies (glasses-free 3D cinema, real-time light field capture and rendering system, full-angle 180 degree light-field display).*

*Patrik Goorts is a postdoctoral researcher at Hasselt University, Belgium, specialized in free viewpoint interpolation and depth estimation.*

*Lode Jorissen is a Ph.D candidate in the Expertise Centre for Digital Media (EDM) at Hasselt University, Belgium. He previously worked on 360 degree video and currently focuses his work on view interpolation using light fields.*

*Adrian Munteanu is professor at Vrije Universiteit Brussel, Belgium. His research interests include image, video and 3D graphics compression, error-resilient coding and multimedia transmission over networks. He is the author of more than 250 journal and conference publications, book chapters and contributions to standards, and received several awards for his work. Adrian Munteanu currently serves as Associate Editor for IEEE Transactions on Multimedia.*

*Beerend Ceulemans is a Ph.D candidate in the Department of Electronics and Informatics (ETRO) at the Vrije Universiteit Brussel (VUB). His research interest are centered on virtual viewpoint synthesis for autostereoscopic 3D screens and free viewpoint video.*

*Pablo Carballeira received the Ph.D. degree in Telecommunication Engineering from the Universidad Politécnica de Madrid (UPM) in 2014. He is with the Grupo de Tratamiento de Imágenes at UPM since 2007 and his current research interests include coding and subjective evaluation of Super Multiview and Free Navigation Video.*

*Sergio García is a Ph.D candidate in the Grupo de Tratamiento de Imágenes (GTI) at the Universidad Politécnica de Madrid (UPM), where he has been working since 2013. His research interests include adaptive streaming techniques and algorithms, as well as 3D graphics compression and rendering, especially in the field of point-cloud-based models.*

*Masayuki Tanimoto received the B.E., M.E., and Dr.E. degrees from the University of Tokyo. He was Professor at Nagoya University and developed FTV (Free-viewpoint Television). Currently, he is Emeritus Professor at Nagoya University and Senior Research Fellow at Nagoya Industrial Science Research Institute. He is Honorary Member of the ITE, Fellow of the IEICE and IEEE Life Fellow. He is chair of the MPEG-FTV group.*