INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2014/M35842
February 2015, Geneva, Switzerland

Source **Poznań University of Technology**
Chair of Multimedia Telecommunications and Microelectronics
Status **Contribution**
Title **Comments on further standardization for free-viewpoint television**
Author Marek Domański (domanski@et.put.poznan.pl),
Adrian Dziembowski,
Krzysztof Klimaszewski,
Adam Łuczak,
Dawid Mieloch,
Olgierd Stankiewicz,
Krzysztof Wegner

1 Introduction

Obviously, the free-viewpoint television (FTV) [1], or broader free-viewpoint video, needs to be standardized as any other communication service. Similarly to other communication services, standardization is rigorously needed for aspects related to the interoperation of hardware and software delivered by various manufacturers and/or placed in various locations. The promising recent research results, visions of innovative companies and substantial technical progress yield significant interests in prospective standardization of FTV [2]. The respective standardization issues are currently explored by Free-Viewpoint Television AHGroup of MPEG. This paper is prepared as a part of this exploration activity.

In this paper we will focus on standardization concerning FTV with arbitrary camera locations around a scene [3,4]. For the sake of simplicity such an arrangement can be roughly approximated by a simple nearly-circular arrangement of cameras. For practical reasons the number of cameras should be limited. This number may be as low as 10 camera for broadcasts of small scenes, maybe wrestling, but may be much higher, maybe 50, for basketball broadcasts. Here, the broadcasts from large playfields, like football or soccer, are left beyond the scope of the considerations.

First of all, we consider the systems where the viewer virtually navigates in a scene by choosing by himself/herself the trajectory of virtual viewpoints. For viewing the synthesized video virtually taken from an arbitrary selected viewpoint, the display may be of various type: monoscopic, stereoscopic, autostereoscopic or three-dimensional with light-field synthesis.

When considering standardization, we have in mind such prospective FTV applications, like e.g. sport broadcasts (judo, boxing, wrestling, sumo, dance etc. but also volleyball or basketball), performances (theater, circus), interactive courses (medical, cosmetics, dance etc.), interactive manuals and school teaching materials.

Obviously, standardization on FTV should include various aspects: multiview video compression, camera system parameter coding, depth data compression, 3D scene representation, spatial audio compression, interactive communication between a user terminal and a server etc. Among the abovementioned issues, video coding is probably the most challenging. Therefore, at first we will review the state-of-the-art 3D video coding technology as a possible candidate for free-navigation applications.

2 MV-HEVC and 3D-HEVC in FTV applications

Here, we going to consider the compression performance of MV-HEVC [5] and 3D-HEVC [6], i.e. the two existing state-of-the-art video compression technologies that are under standardization provided jointly by ISO/IEC and ITU-T. Both technologies have been developed for rather densely spaced cameras, and they have been tested predominantly for linear arrangements of cameras. As mentioned before, here we focus on applications where cameras are located around a scene. The circular camera setup (Fig. 1) is approximated by nearly-circular camera arrangement. Such camera arrangement was used to produce FTV test multiview sequences “Poznan Blocks” and “Poznan Team” (Figs. 2 and 3) [3,7]. For the indoor sequence “Poznan Blocks” the radius of the camera arrangement was R = 3 meters, while for the outdoor test sequence “Poznan Team” the radius was R = 15 meters. These two sequences correspond to realistic scenarios of FTV with scenes of limited dimensions. Even for such limited-size scenes, the distances between cameras were about 60 and 180 centimeters, i.e. the distances are huge as compared to the scenarios used for testing the techniques of MV-HEVC and 3D-HEVC.

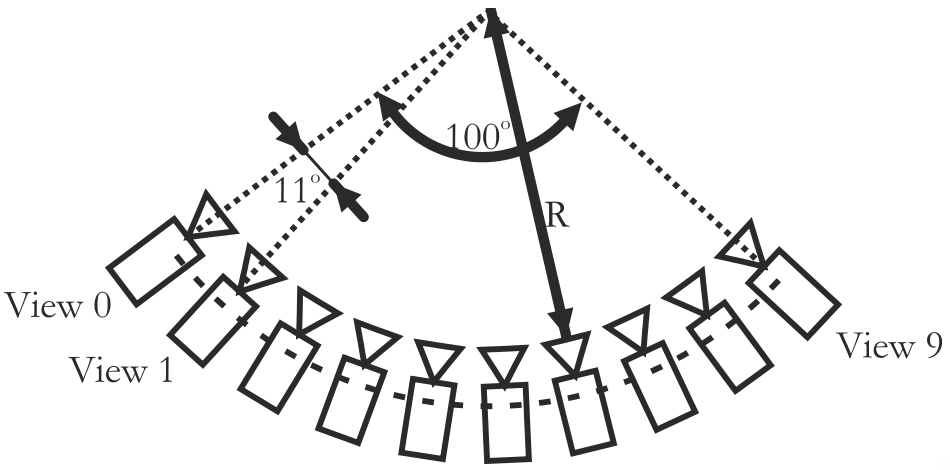

Figure 1. Circular camera setup used in the experiments reported in this document.

Figure 2. Selected frames from “Poznan Blocks” sequence.

Figure 3. Selected frames from “Poznan Team” sequence.

In order to asses available compression efficiency for the sequences with circular camera arrangement three state-of-the-art HEVC-based codecs were used to compress the two abovementioned test sequences. The three codecs are: HEVC simulcast, MV-HEVC and 3D-HEVC. The minimum compression performance is for HEVC simulcast that does not exploit any redundancy related to the view similarity. The question we ask is how much MV-HEVC and 3D-HEVC improve compression efficiency with respect to HEVC simulcast for multiview video obtained from sparely spaced cameras located on an arch.

In the experiments, the view bitrate and view quality was considered only, i.e. no depth data is included into bitrate numbers. The data are provided for 3 views corresponding to cameras No. 4, 5 and 6. The configuration parameters for all HEVC-based encoders were the same: intra-period = 24, B_0 picture period = 8, 1 slice per picture, SAO and VSO switched off. For MV-HEVC and 3D-HEVC, the software version HTM-11.0 was used.

The respective rate-distortion curves are depicted in Fig. 4 for both test sequences "Poznan Blocks" and "Poznan Team". The quality is expressed as average luma PSNR for three views, and bitrate is calculated for three views together. The respective data are collected in Table 1.

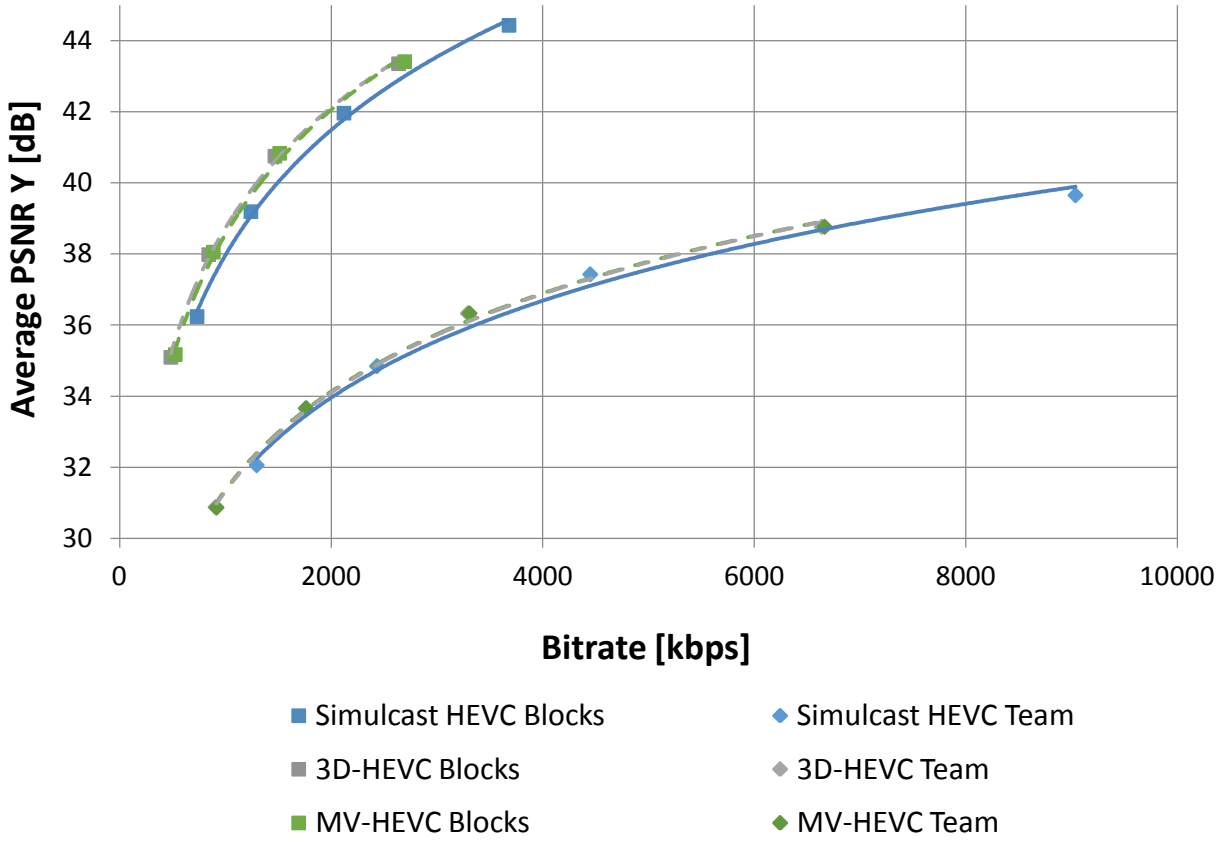

Figure 4. Rate-distortion curves for HEVC simulcast, MV-HEVC and 3D-HEVC for both test sequences "Poznan Blocks" and "Poznan Team".

Table 1. Average luma PSNR and bitrates for 3 views for HEVC simulcast, MV-HEVC and 3D-HEVC for both test sequences “Poznan Blocks” and “Poznan Team”.

HEVC		MV-HEVC		3D-HEVC	
Bitrate [kbps]	PSNR [dB]	Bitrate [kbps]	PSNR [dB]	Bitrate [kbps]	PSNR [dB]
Poznan Blocks					
3682	44.43	2695	43.41	2637	43.35
2121	41.96	1513	40.83	1468	40.75
1241	39.19	886	38.05	842	37.98
732	36.24	526	35.17	484	35.09
Poznań Team					
9038	39.65	6666	38.76	6638	38.76
4450	37.43	3308	36.33	3292	36.33
2433	34.84	1764	33.66	1758	33.66
1296	32.06	916	30.86	907	30.88

The results of the experiments demonstrate, that the specialized MV-HEVC and 3D-HEVC codecs provide only a small improvement over HEVC simulcast, i.e. bitrate reduction with respect to HEVC is very limited. For “Poznan Blocks” the average bitrate reduction is 10.9 % for MV-HEVC, and 13.1% for 3D-HEVC. Nevertheless, for “Poznan Team” with even more sparsely spaced cameras the bitrate reduction is 3.1% for MV-HEVC and 3.6% for 3D-HEVC, again with respect to HEVC simulcast.

The abovementioned observations confirm the straightforward expectation that the compression efficiency gain versus HEVC simulcast decreases as the distance between neighboring cameras increases.

As the bitrate reduction obtained by the use of MV-HEVC and 3D-HEVC versus HEVC simulcast is very limited for FTV video, we probably **need a new compression technology capable to efficiently compress free-viewpoint video that provides the feature of free navigation**.

3 Overview of free-navigation systems

Consider a practical free-viewpoint television system that provides the feature of free navigation a viewer. In principle the system consists of 4 basic units:

- The content acquisition system (cameras , microphones, depth camera, potentially light-field cameras);
- The representation server where system calibration calculations, video and audio preprocessing and 3D scene representation estimation are implemented;
- The rendering server where virtual views and the corresponding audio are synthesized according to the requests of the viewers;

- The user terminal where requested views are presenting together with the corresponding audio. The terminal is capable for bidirectional communications thus allowing the view requests to be transmitted in the uplink.

The block diagram of an FTV system is depicted in Fig. 5.

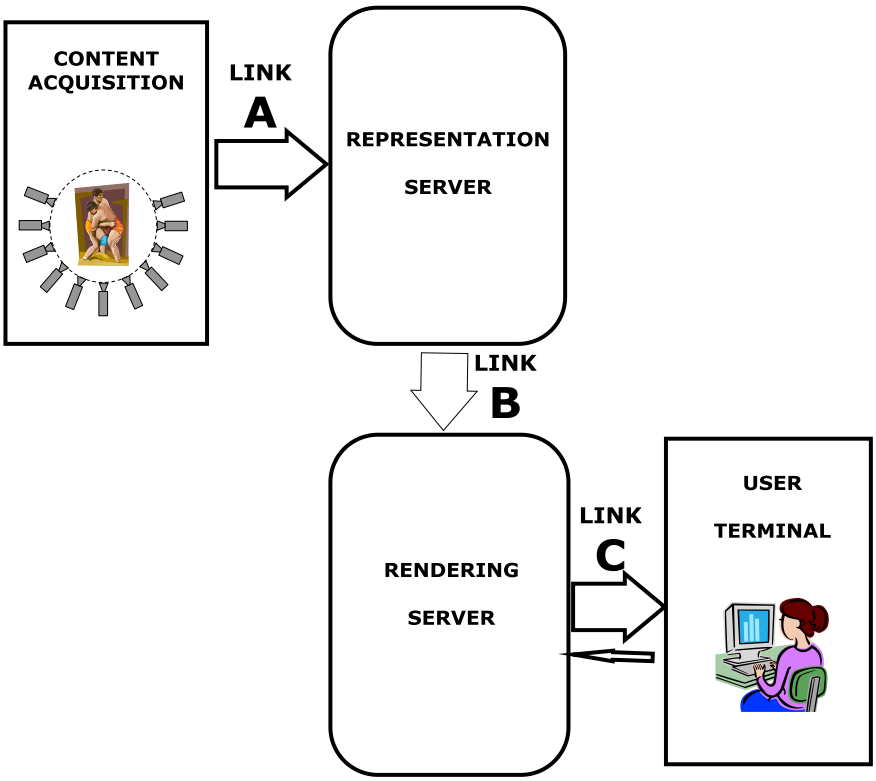

Figure 5. The block diagram of an FTV system with free navigation capability.

The basic functional blocks of an FTV systems are linked by three links named as A, B and C. Some links may vanish in particular configurations of the system. For example, the representation may be calculated directly on site of video and audio acquisition. In such a case, the link A does not exist. The rendering server may be incorporated into the user terminal. Then the link C is superfluous.

The characteristics of the individual links are collected in Table 2.

Table 2. The link characteristics.

Link	Data transmitted	Link features
A	Multiview video with calibration data, Data from depth sensors, Audio from multiple microphones, Output video from light-field cameras.	The link belongs to contribution environment. High quality required. For some simple applications: off-line transportation of media (e.g. HDDs) possible.
B	Audiovisual representation of 3D scene audiovisual representation. *Simple example: Corrected multiview video with depths and spatial audio*.	Broadcasting quality required.
C	Bidirectional link. *Downlink*: Video and audio corresponding to the selected virtual view, possibly (?) also to the neighboring views. Video may be: monoscopic, stereoscopic, multiview for autostereoscopic displays or three-dimensional with light-field synthesis for 3D displays. *Uplink*: Control data representing the free-navigation requests.	Broadcasting quality required.

The vast majority of data is related to video while audio and system parameters probably refer to less than 10% of the data, similarly as in classical television. Therefore we focus on video compression leaving audio for further studies.

Nevertheless, audio compression transmission and processing is an important component of the system. It needs separate studies and will be left beyond the scope of this paper.

Now, we are going to estimate standardization needs with respect to video in the above considered system.

4 Standardization needed for FTV with free navigation

4.1. Link A

For short-term development, maybe either delivery of the materials on digital media (like HDD or SSD) or simulcast transmission may be acceptable. For long-term evolution FTV systems efficient high-fidelity multiview video compression will be needed for sparsely spaced and arbitrary located cameras. As this link belongs to production environment, the compression has to output video that may be subject of edition, therefore the content should be resistant to multiple encoding-decoding cycles.

4.2. Link B

This link may not exist in many prospective systems. If it exists and the servers are distant, for long-term evolution FTV systems, a compression technology is needed to efficiently represent the 3D model of a scene. Unfortunately, the structure of the future 3D scene models is not defined yet. Nevertheless, multiview plus depth representation is quite likely. In such a highly probable case, again the high-fidelity multiview video compression technology is needed for sparsely spaced and arbitrary located cameras. Nevertheless, here the compression technology may use depth and may exploit that the views are corrected.

In the case where the rendering server is a part of user terminal the efficiency of this compression will be critical for practical applications. The application scenario with virtual view synthesis will be killed if an appropriate compression technology will be not developed.

4.3. Link C

In the simple cases when the virtual views are presented on a monoscopic, stereoscopic or autostereoscopic displays, the existing standardized compression technologies like MV-HEVC and 3D-HEVC will be sufficient.

Nevertheless, the standardization in FTV should be aware of development of 3D displays. Very realistic free navigation will be possible with true 3D displays that synthesize light fields. For such displays, probably the efficient compression technology will be needed for sparsely spaced and located on an arch cameras. Perhaps, only small regions of some views need to be transmitted (e.g. disoccluded regions [8]). The remaining virtual views will rendered within the user terminal.

Despite of the compression standardization, the bidirectional communication must be standardized in order to allow interoperability for free navigation.

5 Conclusions

The above considerations lead to a conclusion that an efficient multiview/3D video compression technology is needed for sparsely spaced and arbitrary located cameras. This technology would be probably needed in two versions:

a) High-fidelity contribution profile: no depth available for compression of video, high-quality decoded video resistant to multiple decoding-encoding cycles (for link A).

b) Consumer profile: depth available for compression of video, broadcast quality (links B and C, the latter for 3D displays only).

The prospective standardization works should be aware of the basic compression progress (see Fig. 6). It would be advantageous if the requested new compression technology will be transparent to the single-view compression technology. It may be expected that the new generation of single-view compression will be available when the final new 3D video compression technology will be finalized.

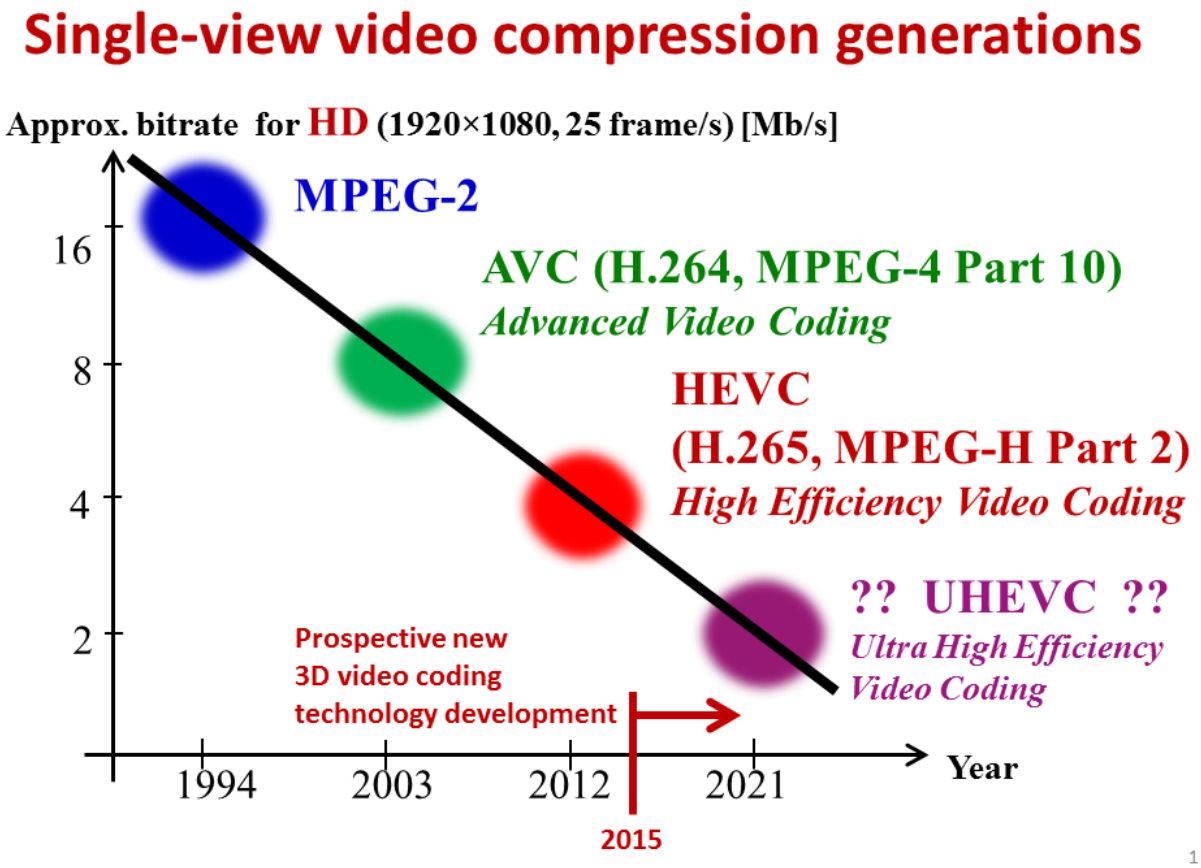

Figure 6. New 3D video standardization activity versus expected single-view video compression development.

6 Acknowledgement

The work was supported by National Science Centre, Poland, according to the decision DEC-2012/05/B/ST7/01279.

References

[1] M. Tanimoto, M. Tehrani, T. Fujii, T. Yendo, “FTV for 3-D spatial communication”, Proc. IEEE, Vol. 100, pp. 905-917, April 2012.

[2] M. Tanimoto, T. Senoh, S. Naito, S. Shimizu, H. Horimai, M. Domański, A. Vetro, M. Preda, K. Mueller, Proposal on a new activity for the third phase of FTV, ISO/IEC JTC 1/SC 29/WG 11, Doc. M30229/M30232, Vienna, Austria, July/Aug. 2013.

[3] M. Domański, A. Dziembowski, A. Kuehn, M. Kurc, A. Łuczak, D.Mieloch, J. Siast, O. Stankiewicz, K. Wegner, “Experiments on acquisition and processing of video for free-viewpoint television”,3DTV-CON, Budapest, July 2014.

[4] M. Domański, “Practicing free-viewpoint television: multiview video capture and processing,” in: M. Tanimoto, T. Senoh, “FTV seminar report,” ISO/IEC JTC 1/SC 29/WG 11, Doc. MPEG M34564, Sapporo, Japan, July 2014.

[5] G. Tech, K. Wegner, Y. Chen, M. Hannuksela, J. Boyce, “MV-HEVC Draft Text 8”, JCT-3V of ITU-T, ISO/IEC Doc. JTC3V-H1004, 2014.

[6] G. Tech, K. Wegner, Y. Chen, S. Yea, “3D-HEVC Draft Text 4”, JCT-3V of ITU-T and ISO/IEC Doc. JTC3VH1001, 2014.

[7] M. Domański, A. Dziembowski, M. Kurc, A. Łuczak, D. Mieloch, J. Siast, O. Stankiewicz, K. Wegner, "Poznan University of Technology test multiview video sequences acquired with circular camera arrangement – “Poznan Team” and “Poznan Blocks” sequences " ISO/IEC JTC1/SC29/WG11 MPEG2014/M35846, Geneve, Switzerland, February 2015.

[8] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, „Technical Desciption of Poznan University of Technology proposal for Call on 3D Video Coding Technology”, ISO/IEC JTC1/SC29/WG11 MPEG2011/M22697, Geneve, Szwajcaria, November 2011.