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/ m38101
February 2016. San Diego. US

**Source** **Poznań University of Technology**
**Chair of Multimedia Telecommunications and Microelectronics**

**Status** **Contribution to the 114th MPEG Meeting**

**Title** **[FTV AHG] 3D-HEVC extensions for free navigation**

**Author** **Marek Domański (domanski@et.put.poznan.pl).**
**Jarosław Samelak (jsamelak@multimedia.edu.pl).**
**Olgierd Stankiewicz (ostank@multimedia.edu.pl).**
**Jakub Stankowski (jstankowski@multimedia.edu.pl).**
**Krzysztof Wegner (kwegner@multimedia.edu.pl).**

# 1. Introduction

The free viewpoint television (FTV) [1] is a prospective interactive internet video service that will provide a viewer the ability to navigate virtually around a scene. The development of efficient. reliable and high-quality FTV systems is still a demanding challenge for the research and the technology [2]. and further also for standardization. The MPEG (ISO/IEC JTC1/SC29/WG11) group of experts has already recognized this challenge by issuing the Call for Evidence on Free-Viewpoint Television [3]. The Call is aimed at identification of a compression technology that would be more efficient than the existing 3D-HEVC technology. in particular. for applications oriented towards free virtual navigation and super multiview video.

In response to this Call. among the others. the document m37893 [4] was submitted. The document m37893 reports a technique that outperforms standard 3D-HEVC for the multiview plus depth video obtained from cameras located on an arc. The document provides only a brief description of the tools added or modified in the 3D-HEVC software [5]. according to the requirements of the Call.

Here. in this document we provide additional augmenting and wider information that enhances the document m37893.m.

## 2. Prospective standardization for FTV

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. sports coach study materials and school teaching materials. and others. Practical systems for such applications should provide the feature of free navigation a viewer. A system that provides the ability of free virtual navigation 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. 1 [5].

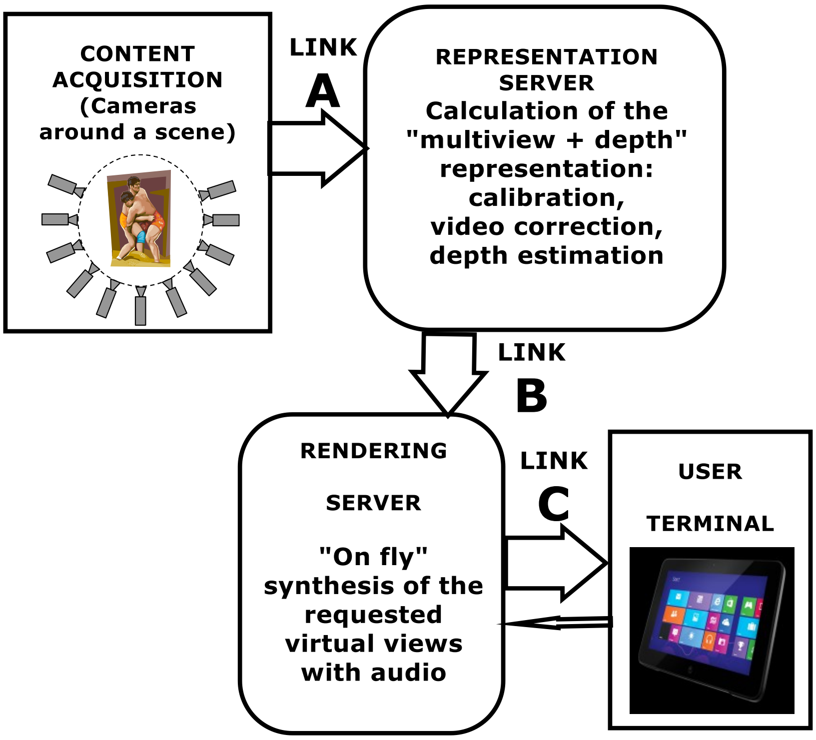

Fig. 1. The block diagram of the systems providing the ability of free navigation.

Prospective standardization on FTV may 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.

As already discussed elsewhere [6-8]. among the video compression problems. the need for an efficient compression technology for the link B is the most urgent and demanding .

It might appear astonishing. that the recently standardized MV-HEVC and 3D-HEVC techniques are insufficient the FTV and free navigation applications as mentioned already in [6]. The MV-HEVC and 3D-HEVC technologies have been tested and optimized for multiview video acquired using a set of cameras densely located on a line. This is not a practical scenario for a natural content needed for free navigation. In such applications. the camera will be sparely distributed around a scene. A model of such camera distribution is a camera arrangement where camera are placed on an arc and the angles between the neighboring cameras are significant like in the test sequences “Poznan Blocks” and “Poznan Team” [6] where the angle between the neighboring convergent camera axes is around 10 degrees. For such camera arrangements. the standard MV-HEVC and 3D-HEVC technologies provide compression efficiency not much higher than that obtained by independent compression of each view by the use of the standard mono-view HEVC compression (see Fig. 2 [6]).

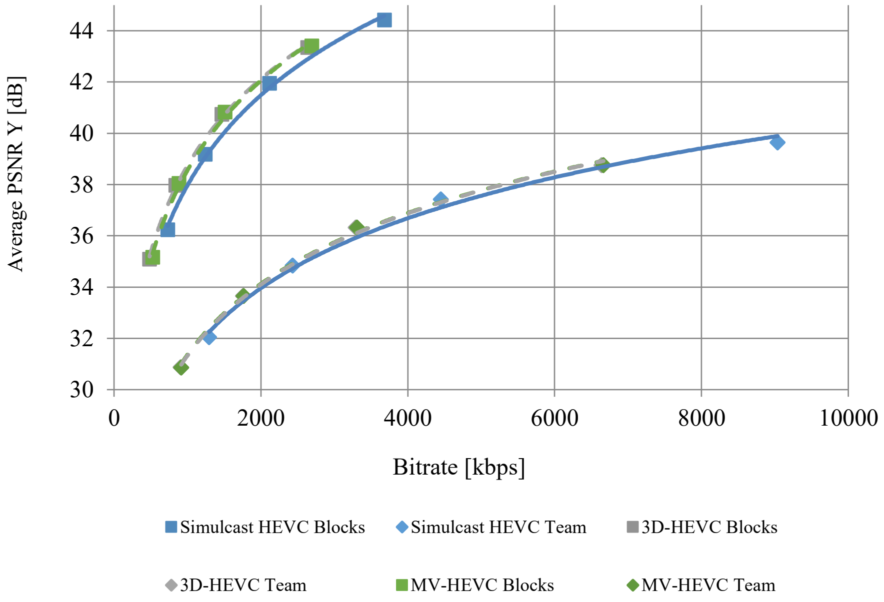

Fig. 2. Compression of “Poznan Blocks” and “Poznan Team” sequences (the first 50 frames) for 3 views (reproduced from [6]).

## 3. Proposed 3D-HEVC extension for arbitrary camera locations

In principle. the inter-view prediction which is the core tool of MV-HEVC. in general. does not assume any particular view arrangement. But. in order to increase the coding efficiency of 3D-HEVC. a number of specific tools was introduced and others were simplified with explicit 1D parallel view arrangement assumption.

Repeated the reasoning from [7] we admit that a significant simplification was used in 3D-HEVC that states that the views are vertically aligned. Therefore. disparity vectors are restricted only to horizontal direction. and they can be derived from depth data through simple linear equation.

$$\begin{aligned} d_h &= a \cdot v + b \\ d_v &= 0 \end{aligned} \qquad (1)$$

where the horizontal component $d_h$ of the disparity vector is calculated from depth sample value $v$ and scaled to disparity by using scale factor $a$ and offset $b$ transmitted for each view. and vertical component of disparity vector $d_v$ is always set equal to zero.

In the proposed 3D-HEVC extension [7]. we remove the restriction that the disparity has a nonzero component for horizontal direction only. Instead. we apply general derivation of disparity vector based on depth data.

From the point position $x.\ y$ in the picture and associated depth sample value $v$. position of the corresponding point $x_r.\ y_r$ in the reference view can be calculated. as already shown in paper [7]. The derived disparity vector is simply the difference in the position of considered point and the corresponding point in the reference view

$$\begin{aligned} d_h &= x_r - x \\ d_v &= y_r - y \end{aligned} \qquad (2)$$

The position of the corresponding point in the reference view can be derived through depth based projection of the point according to the equation (3)

$$\begin{bmatrix} z_r \cdot x_r \\ z_r \cdot y_r \\ z_r \\ 1 \end{bmatrix} = P_r \cdot P^{-1} \cdot \begin{bmatrix} z \cdot x \\ z \cdot y \\ z \\ 1 \end{bmatrix}. \qquad (3)$$

where $P_r$. and $P$ are projection matrices for both the reference view and the view being coded. We have defined projection matrices as a multiplication of intrinsic and extrinsic camera parameters that need to be transmitted for each view being coded in the bitstream.

$$\boldsymbol{P} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{R} & -\boldsymbol{R} \cdot \boldsymbol{T} \\ \boldsymbol{O}^T & 1 \end{bmatrix} \qquad (4)$$

Necessary distance $z$ of the considered point from camera of interest can be calculated [7] based on the depth sample value $v$:

$$z = \left( \frac{v}{2^{DepthMapBitDepth}} \cdot \left( \frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}} \right)^{-1} \qquad (5)$$

where $Z_{near}$ and $Z_{far}$ are depth maps normalization parameters transmitted in the bitstream. and $DepthMapBitDepth$ is bit depth of the depth sample value (typically 8 or 16 bits per sample is used).

Actual camera parameters for each view are a natural extension of the simplified camera parameters already used in 3D-HEVC. Beside already transmitted parameters: focal length $f_x$. optical center $c_x$ of the camera and position of the camera optical center along x axis $T_x$. we have to transmit additional intrinsic parameters: second focal length along vertical direction $f_y$. position of the optical center along second axis $c_y$. and extrinsic parameters: rotation matrix $\boldsymbol{R}$ and remaining coordinates of camera position $T_y$ and $T_z$. $T_y$ and $T_z$ together with $T_x$ create vector $\boldsymbol{T}$.

In the proposed 3D-HEVC extension we have modified disparity vector derivation process in such tools as Disparity Compensated Prediction (DCP). Neighboring Block Disparity Vector (NBDV). Depth oriented NBDV (DoNBDV). View Synthesis Prediction (VSP). Inter-view Motion Prediction (IvMP). Illumination Compensation (IC) [7].

In the 3D-HEVC. contrary to the proposed extension. derived disparity vectors do not depend on the position in view being coded. For example. in the DoNBDV tool. 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.

Considering circular (arc) view arrangements. a given depth sample value can represent a different distance plane for each view (which is not the case in 1D parallel view arrangement). In the proposed extension. corresponding depth sample value in the reference view can be derived simultaneously with the disparity vector through the equation (3). The equation (3) provides distance $z_r$ of the correspondent point from the reference camera which can be normalized to depth sample value with the help of

$$v = \frac{\frac{1}{z} - \frac{1}{z_{far}}}{\frac{1}{z_{near}} - \frac{1}{z_{far}}} \cdot 2^{DepthMapBitDepth} \qquad (6)$$

The derivation of depth sample value has been applied to the synthesis of dependent views' depth maps.

## 4. Experimental results and conclusions

The experimental results are summarized in Table 1. These results demonstrate some improvement of the compression efficiency for the free navigation content in the Link B, where many views need to be transmitted. The data are provided for the available test sequences with the circular camera arrangements [9-11].

Table 1 . Summary of the experimental results for the test sequences with circular camera arrangements. The results are provided for 7 views.

| | Our vs HEVC simulcast | 3D-HEVC vs simulcast | Our vs 3D-HEVC | Our vs MV-HEVC | 3D-HEVC vs MV-HEVC |
|---|---|---|---|---|---|
| Poznan Blocks | -19.74% | -14.63% | -6.44% | -4.20% | 2.37% |
| BBB_Flowers | -10.78% | -8.24% | -3.03% | -2.80% | 0.21% |
| Ballet | -33.91% | -28.07% | -8.64% | -12.55% | -4.32% |
| Breakdancers | -40.72% | -34.88% | -9.79% | -13.71% | -4.39% |
| **Average** | **-26.29%** | **-21.46%** | **-6.97%** | **-8.32%** | **-1.53%** |

For the above comparison, for the MV-HEVC and 3D-HEVC implementations the HTM13.0 was used. The proposed technique was implemented on top of HTM13.0.

## 5. Acknowledgement

This research project was supported by public funds in the course of DS "Działalność Statutowa" project.

## 6. References

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

[2] G. Lafruit. .M. Domański. K. Wegner. T. Grajek. .T. Senoh. J. Jung. .P. .Kovács. P. Goorts. L. Jorissen. A. Munteanu . B. Ceulemans. P. Carballeira. S. García. Ma. Tanimoto. “New visual coding exploration in MPEG: Super-MultiView and Free Navigation in Free viewpoint TV”. *IST Electronic Imaging 2016.* February2016. San Francisco. California.

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

[4] M. Domański. A. Dziembowski. A. Grzelka. Ł. Kowalski. D. Mieloch. J. Samelak. O. Stankiewicz. J. Stankowski. K. Wegner. “Technical Description of Poznan University of Technology proposal for Call for Evidence on Free-Viewpoint Television”. *ISO/IEC JTC1/SC29/WG11*. Document MPEG2016/M37893. February 2016. San Diego. USA.

[5] M. Domański. A. Dziembowski. K. Klimaszewski. A. Łuczak. D. Mieloch. O. Stankiewicz. K. Wegner. “Comments on further standardization for free-viewpoint television”. *JTC1/SC29/WG11*. Document MPEG2015/M35842. February 2015. Geneva. USA.

[6] M. Domański. A. Dziembowski. D. Mieloch. A. Łuczak. O. Stankiewicz. K. Wegner. “A practical approach to acquisition and processing of free viewpoint video.” *Picture Coding Symposium PCS 2015*. Cairns. Australia. pp. 10-14.

[7] J. Stankowski. Ł. Kowalski. J. Samelak. M. Domański. T. Grajek. K. Wegner „3D-HEVC extension for circular camera arrangements.” *3DTV-CON*. Lisbon 2015.

[8] M. Domański. M. Bartkowiak. A. Dziembowski. T. Grajek. A. Grzelka. A. Łuczak. D. Mieloch. J. Samelak. O. Stankiewicz. J. Stankowski. K. Wegner. “New Results in Free-viewpoint Television Systems for Horizontal Virtual Navigation”. submitted to *IEEE International Conference on Multimedia and Expo 2016 (ICME)*. Seattle 2016.

[9] M. Domański. et al.. „Poznan Blocks – a multiview video test sequence and camera parameters for FTV”. *MPEG* M32243. San Jose. Jan. 2014.

[10] P. T. Kovacs. “[FTV AHG] Big Buck Bunny light-field test sequences”. *JTC1/SC29/WG11*. Doc. MPEG M35721. Geneva. Feb. 2015.

[11] C.L. Zitnick. et al.. "High-quality video view interpolation using a layered representation." *ACM Trans. Graphics*. vol. 23. pp. 600-608. Aug. 2004.