NEW CODING TECHNOLOGY FOR 3D VIDEO WITH DEPTH MAPS AS PROPOSED FOR STANDARDIZATION WITHIN MPEG

Marek Domański, Tomasz Grajek, Damian Karwowski, Krzysztof Klimaszewski, Jacek Konieczny, Maciej Kurc, Adam Łuczak, Robert Ratajczak, Jakub Siast, Olgierd Stankiewicz, Jakub Stankowski, Krzysztof Wegner

Chair of Multimedia Telecommunications and Microelectronics, Poznań University of Technology,
Email: domanski@et.put.poznan.pl

ABSTRACT

In this paper, we present the original compression technology that was designed and developed at Poznań University of Technology in response to MPEG Call for Proposals on 3D Video Coding Technology. One view is coded in the HEVC syntax while the side views and the depth maps are very efficiently coded exploiting various redundancies that are related to similarities between individual views and depth maps. Even 2 side views and 3 depth maps can be coded at the total bitrate that usually does not exceed 50% of the bitrate needed for a single view.

***Index Terms*—** 3D video, compression, 3D video coding, HEVC, depth coding, CfP, MPEG, MPEG-H

1. INTRODUCTION

Stereoscopic video compression and transmission is becoming a mature technology used in many applications. On the other hand, nowadays, the research is focused on more advanced 3D video systems of the second generation. Such system are aimed at provision of more realistic spatial video to a viewer. For example they are able to reproduce the sense of movement parallax. Currently, the following two prospective applications that attract much attention:

- **Autostereoscopic displays** provide binocular perception of stereoscopic depth without the use of special glasses. Such a device simultaneously displays many views corresponding to the scene seen from slightly different viewing angles. In order to ensure comfortable perception of stereoscopic video, the large autostereoscopic television displays need to produce many views. The state-of-the-art displays produce 28 views, but maybe even more views will be produced by the forthcoming displays in the nearest future.
- **Free viewpoint television** is an interactive service where the viewers are able to navigate their virtual viewpoints and viewing directions for the monoscopic or stereoscopic viewing of the scene. Such a technology needs many views to be available for display. The receiver should be capable to synthesize the additional intermediate views in such a way that the viewpoint and the viewing direction may be changed very smoothly.

Both abovementioned applications require transmission of reach 3D video content that usually also includes depth maps. The sample values in a depth map correspond to the depth, i.e. to the physical distance to the object in the 3D space. More often the sample values in a depth map correspond to the disparity that is uniquely related to the depth. Therefore new compression technology is needed for efficient transmission of the video content consisting of several views and the corresponding depth maps.

Identification of such needs has resulted in the standardization process initiated by MPEG, i.e. an expert group affiliated by International Organisation for Standardisation (ISO). In 2001, MPEG has started FTV (Free-viewpoint TV) project, which is a new framework that allows viewing of a 3D world by freely changing the viewpoint [1]. Multiview Video Coding (MVC) was the first phase of FTV, which enabled the efficient coding of multiple camera views. MVC has been standardized in Annex H to the Advanced Video Coding (AVC) standard [2].

Multiview Video Coding exploits the redundancy that exists between the neighboring views, e.g. between the left and the right views. In that way the bitrate is usually reduced by 15-30% with respect to the simulcast compression of all views.

On the other hand, in video compression, recent ongoing standardization activities are very successful. They have resulted in new generation of video coding technology called High Efficiency Video Coding (HEVC) [3] that provides bitrate reduction of about 40-45% as compared to AVC. It was shown that even simulcast coding of multiple views using HEVC is much more efficient than MVC [4].

Simultaneously the FTV project came to the second phase, which is 3D Video Coding that should enable efficient coding of multiple views together with the corresponding depth maps. Prospective applications include high-quality autostereoscopic displays. Moreover, the new technology is expected to enable stereo devices to cope with varying display types and sizes, and different viewing preferences. This includes the ability to vary the baseline distance for stereo video to adjust the depth perception, which could help to avoid fatigue and other viewing discomforts [1].

In order to define the state-of-the-art 3D video coding technology, MPEG has announced Call for Proposals (CfP) on 3D Video Coding Technology [5]. After longer discussion it was decided to have this CfP in two categories: backward compatible with MVC and compatible with HEVC.

By the end of August 2011, in response to CfP, over 20 proposals have been submitted in the two abovementioned categories. The proposals have been ranked using subjective quality assessment of the decoded video clips. The tests have been done in 13 test laboratories from all the parts of the world. It was one of the very first subjective evaluation experiments devoted to evaluation of compression efficiency in 3D video. It was also a large experiment that involved more than 600 viewing subjects who had to assess more than 2700 3D video clips.

The results of subjective tests were disclosed during MPEG meeting in Geneva in the end of November 2011. In the HEVC-compatible class, the proposals from Fraunhofer Institute – H. Hertz Institute Berlin and from Poznań University of Technology were qualified as the best. Therefore it was decided to merge these two solutions in one software in the shortest time.

In this paper, we present a technical description of the original compression technology designed and developed at Poznań University of Technology in response to Call for Proposals on 3D Video Coding Technology [5].

2. 3D VIDEO CODING TECHNOLOGY

The proposed HEVC-compatible compression technology is intended to encode a limited number of video views together with the corresponding depth maps. It is assumed that other views are synthesized in the receiver if needed for display. The technique called DIBR (Depth-Image-Based Rendering) [6] has been developed for that purpose.

One of the views is coded in the HEVC syntax while in side views only the disoccluded regions are coded and transmitted, and the remaining parts are reconstructed at the receiver from the base view using DIBR [7]. The shapes of disoccluded regions are derived in the decoder from reconstructed depth maps. Therefore no side information on shape needs to be transmitted

Proposed technology exploits high correlation in neighboring views by means of inter-view disparity compensation mechanism similar to the one used in the MVC extension of the AVC, and Depth-Based Motion Prediction (DBMP) [8] which infer motion information form the already encoded views.

The encoder (Fig. 1) is composed of five sub-encoders: HEVC-compatible base-view encoder, encoder for camera parameters, HEVC-based depth and texture encoders (used for coding of disoccluded parts of the side views) and HEVC-based encoder for a so-called residual layer. Similarly, a decoder consists of five sub-decoders (Fig. 2).

Input videos are split into two layers, the texture layer, which contains content that can be efficiently coded with classic predictive coding, and the residual layer, which contains high frequency residual content that can be represented

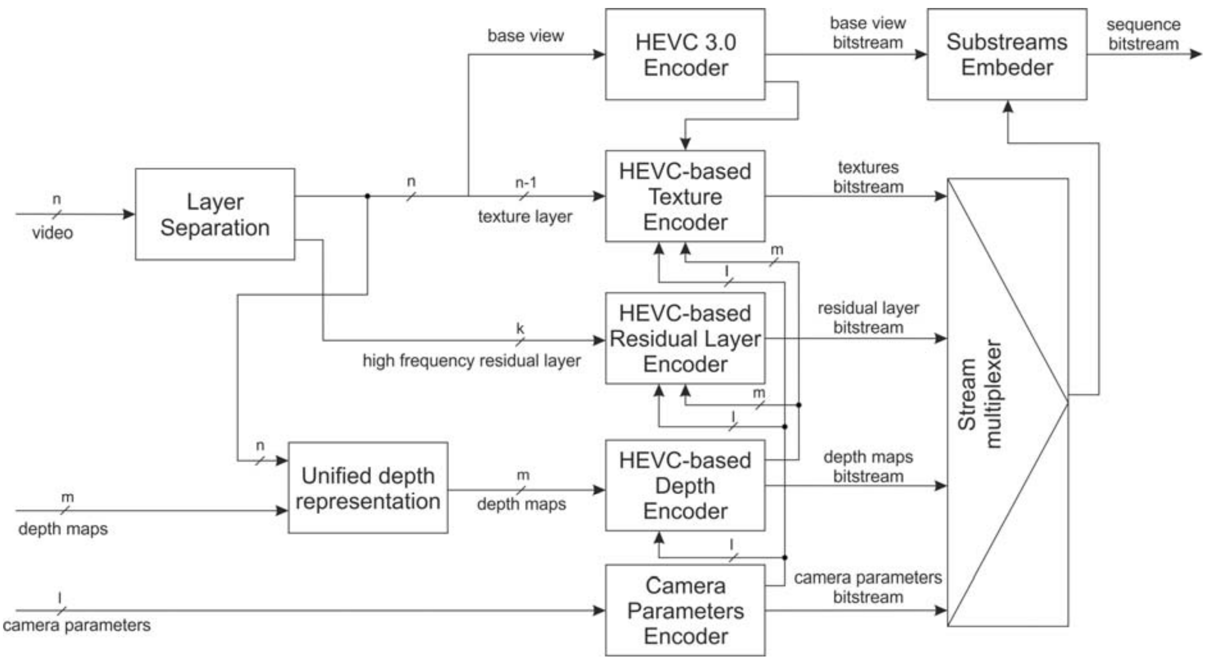

Fig. 1. The block scheme of the encoder.

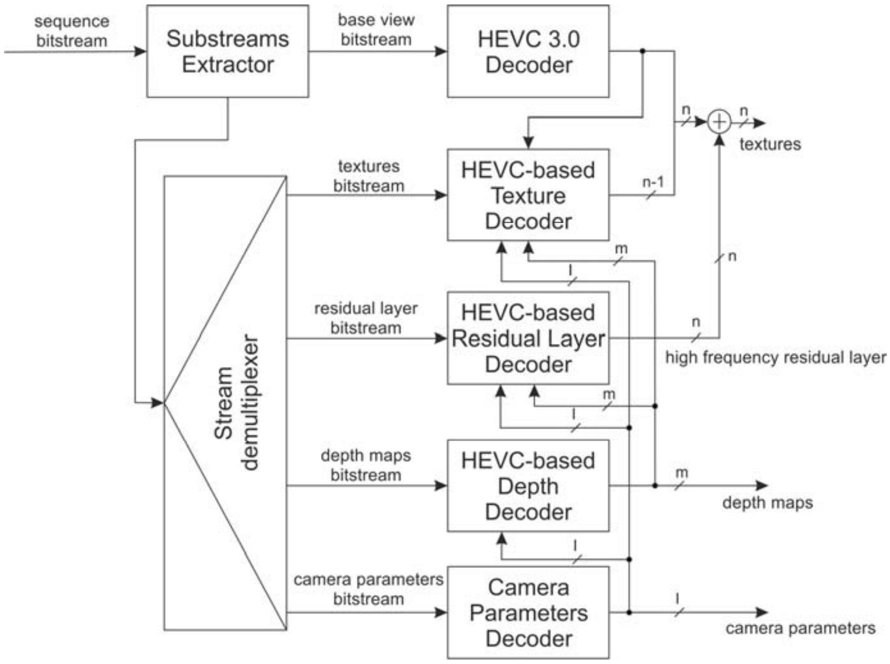

Fig. 2. Scheme of the described decoder.

jointly for several views. Spatial energy distribution of the residual layer is estimated block-wise and compressed by HEVC-based residual layer encoder, while spectral envelope is estimated from energy-normalized residual layer with the use of technique similar to LPC and coded with the use of Log-Area-Ratio representation.

In order to reduce spatial inconsistency in the input depth maps, these depth maps are projected onto the base view and merged together into a unified depth representation, and then re-projected onto the individual views. In that way a refined set of depth maps is produced.

Human perception of depth depends on the distance to the objects. Therefore the near objects are represented more accurately than distant ones. The basic idea is to increase texture quality of objects in the foreground and to increase compression (decrease texture quality) for objects in the background.

3. EXPERIMENTAL RESULTS

During evaluation of the proposals submitted in response to CfP, the individual codecs have been tested in two scenarios:

- transmission of two views and two depth maps,
- transmission of three views and three depth maps.

For each scenario, the quality of the decoded video clips have been assessed in formal subjective testing [5] by independent laboratories. Each proposal was tested by two laboratories that remain undisclosed. For the selected bitrates, the quality of the decoded stereoscopic video was compared to the quality of the decoded video clips obtained from the simulcast HEVC codec. Subjective quality tests were performed using polarization stereoscopic displays as well as with the use of autostereoscopic displays. For the stereoscopic viewing tests, virtual stereo pairs were generated from decoded sequences using DIBR. The Single Stimulus Impairment Scale (SSIS) test method was used with 11 quality levels. All tests were carried out with naive viewers. Mean Opinion Scores (MOS) and confidence intervals corresponding to 95% probability were computed. The results averaged for all 8 test 3D video clips are presented in Figs. 3-5. The quantization steps were chosen in such a way that, at a given “ratepoint”, the quality of the HEVC-decoded clips was the same for various test sequences.

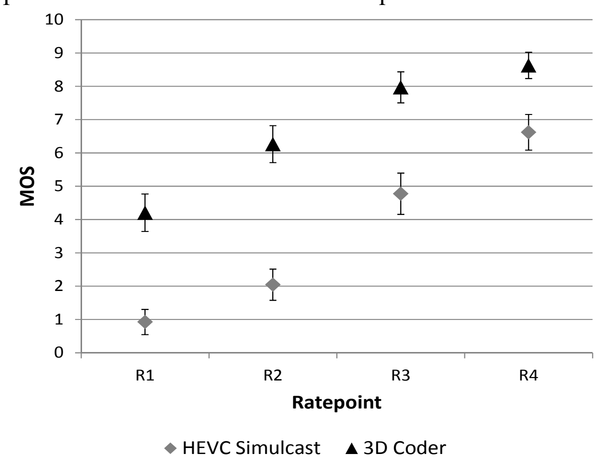

Fig. 3. The viewing results for the 2-view case.

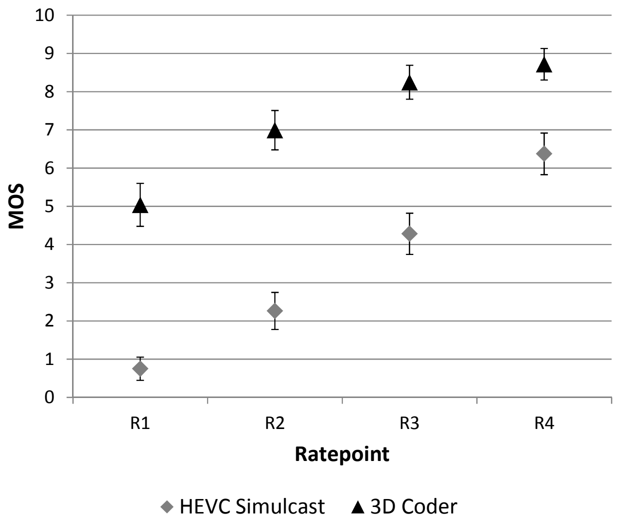

Fig. 4. The viewing results for the 3-view case.

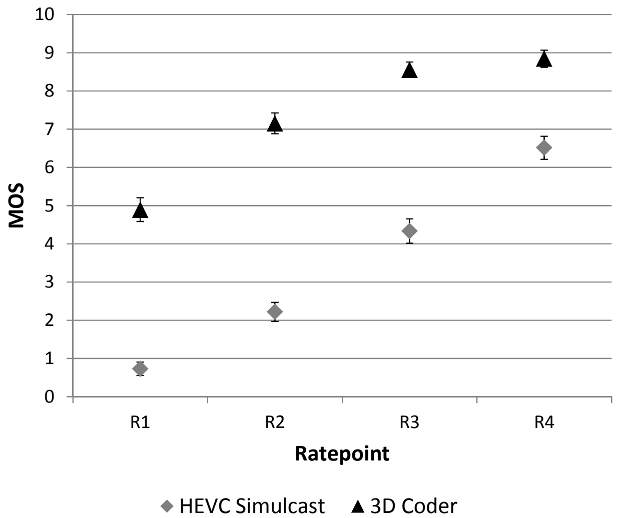

Fig. 5. MOS values obtained
by the use of an autostereoscopic display (28 views).

4. CONCLUSIONS

The original 3D video compression technology has been briefly described. The responses to MPEG Call for Proposals on 3D Video Coding Technology have proven, that this technology may be considered as the-state-of-the-art technology because all other submitted codecs proved either similar performance, or mostly worse performance.

Our codec exploits various redundancies that are related to similarities between individual views and depth maps. In that way, the side views and the depth maps may be represented by very low number of bits leaving most of the bitrate to the base view. The bitrate spent for the base view is in average 65-75% of the total bitrate, depending of the number of side views (2 or 3). The base view is coded in the HEVC syntax while the side views as well as depth data are very efficiently coded using the original techniques designed and developed by the authors. Even 2 views and 3 depth maps are codable at the total bitrate that usually does not exceed 50% of the bitrate needed for a single view. The experimental results prove that 2-3 HD video views of broadcasting quality together with the respective depth maps may be transmitted at the bitrates around 6 Mbps. In comparison to the HEVC simulcast compression, the bitrate reduction exceeding 50% is achieved. In comparison to MVC coding, the bitrate reduction is usually significantly higher. For the proposed codec, the results of the subjective quality assessment (see Figs. 3-5) have been obtained by two independent laboratories. The names of the laboratories that have tested this proposal remain undisclosed. Moreover the laboratory staff was also not informed about the proponent name. Therefore these subjective tests results should be considered as very reliable measures of the codec performance.

Acknowledgement: This work was supported by the public funds as a research project.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 (MPEG), Doc. N12035, “Applications and Requirements on 3D Video Coding”, Geneva, Switzerland, March 2011.

[2] ISO/IEC 14496-10, International Standard, “Generic coding of audio-visual objects – Part 10: Advanced Video Coding,” 6nd Ed., 2010, [also:] ITU-T Rec. H.264, Edition 5.0 (version 11), 2010.

[3] Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-F803, B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, T. Wiegand (Editors), “WD4: Working Draft 4 of High-Efficiency Video Coding”, Torino, Italy, July 2011.

[4] K. Wegner, O. Stankiewicz, K. Klimaszewski, M. Domański, ”Comparison of multiview compression performance using MPEG-4 MVC and prospective HVC technology”, ISO/IEC JTC1/SC29/WG11 (MPEG), Doc. M17913, Geneva, Switzerland, July 2010.

[5] ISO/IEC JTC1/SC29/WG11 (MPEG), Doc. N12036, "Call for Proposals on 3D Video Coding Technology", Geneva, Switzerland, March 2011.

[6] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, M. Tanimoto, "View generation with 3D warping using depth information for FTV". Signal Processing: Image Communication, Vol. 24. Issue 1-2, Pages 65-72, 2009.

[7] 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 description of Poznań University of Technology proposal for Call on 3D Video Coding Technology”, ISO/IEC JTC1/SC29/WG11 (MPEG), Doc. M22697, Geneva, Switzerland, November 2011.

[8] J. Konieczny, M. Domański, “Extended inter-view direct mode for multiview video coding”, IEEE Int. Conf. Acoustics Speech Signal Proc., Prague, Czech Republic, May 2011.