3D VIDEO COMPRESSION BY CODING OF DISOCCLUDED REGIONS

Marek Domański, Jacek Konieczny, Maciej Kurc, Robert Ratajczak,
Jakub Siast, Olgierd Stankiewicz, Jakub Stankowski, Krzysztof Wegner

Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology, Poznań, Poland

ABSTRACT

In this paper, we present the most efficient coding tools that are used in a new video compression technology for multiple views with the depth maps. This very well performing technology was designed and developed in response to MPEG Call for Proposals on 3D Video Coding Technology. The proposed technology exploits a tool that reduces the side views and the side depth maps to small disoccluded regions. This way only one central view and one depth map are coded in the HEVC syntax while the remaining views and the depth maps are synthesized in the decoder from the small disoccluded regions and from the central-view data. Therefore, the bitrate needed for a side view is mostly below 20% of the bitrate for single-view video.

Index Terms— 3D video, compression, HEVC, MPEG

1. INTRODUCTION

Recently, advanced 3D video systems are subject to extensive research. Such systems are able to provide realistic impression of 3D including the movement parallax. They should also provide 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. Prospective applications include glassless autostereoscopic displays and free-viewpoint television that is an interactive service where the viewers are able to change smoothly the viewpoint and the viewing direction.

The abovementioned applications require transmission of 3D video content that may include several views and also depth maps. The sample values in a depth map correspond to the depth values, i.e. to the physical distances to the objects in the 3D space. More often the sample values in a depth map correspond to the disparity values that are related to the depth values by a simple algebraic equation.

Transmission of several views and the corresponding depth maps requires compression technology that is able to exploit efficiently the information redundancy related to the existence of multiple views and depth maps in the video taken for a scene. Therefore, in 2001, Moving Picture Experts Group (MPEG) has started the FTV (Free-viewpoint TV) standardization project [1]. Multiview Video Coding (MVC) was the first phase of this project. MVC has been standardized as a part of the Advanced Video Coding (AVC) standard [2]. This technology enables efficient coding of video acquired from multiple cameras. It exploits the redundancy that exists between the neighboring views, e.g. between the left view and the right view in stereoscopic video. In that way the bitrate is usually reduced by 15-30% with respect to the simulcast compression of all views.

The second phase of the FTV standardization project is 3D Video Coding that should enable efficient coding of multiple views with the depth maps. In order to define efficient coding technology, MPEG has announced Call for Proposals (CfP) on 3D Video Coding Technology [3]. In August 2011, over 20 proposals have been submitted for testing. According to the results that were disclosed during 98th MPEG meeting in Geneva in November 2011, the proposal from Poznań University of Technology [4] was qualified as one of the best performing. Below, the basic coding tools of this technology are described.

This technology is based on the new generic video coding technology called High Efficiency Video Coding (HEVC) [5] 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 clearly more efficient than specialized MVC [6]. In the proposed method, usually one view is coded in the HEVC syntax while the other views and the depth maps are very efficiently coded exploiting the mutual dependencies between individual views and depth maps. The technology is appropriate for compression of arbitrary number of views and corresponding depth maps. Nevertheless, for the sake of simplicity, it will be described under the assumption that 3 views are transmitted: the central view and the side views. This technology comprises many new coding tools but the paper will address only those that have the major impact on compression performance.

2. NON-LINEAR DISPARITY REPRESENTATION

The straightforward approach to disparity map transmission would use uniformly quantized disparity values. Unfortunately such representation does not match the properties of the human visual system that is more tolerant to disparity errors in the background of a scene. Therefore we propose to use non-uniform quantization, so that distant

Figure 1. Original frame from Poznan Street sequence (left view)

Figure 2. Disoccluded regions (white) in Poznan Street sequence

Figure 3. Coded regions (white) in Poznan Street sequence

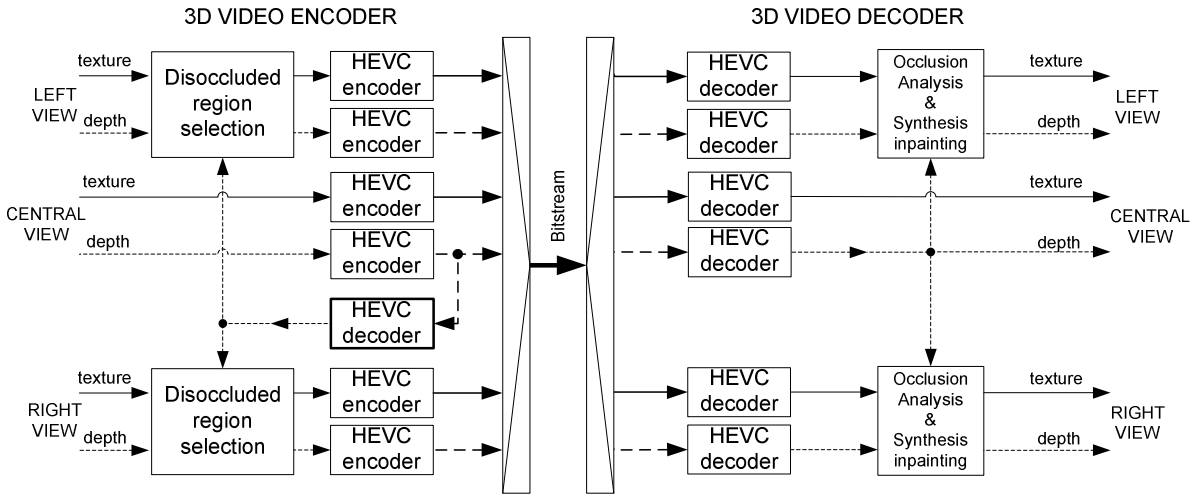

Figure 4. The block diagram of a 3D video codec based on coding of disoccluded regions in the side views

objects are quantized more roughly than the closer ones. In order to be compliant with HEVC quantization schemes we propose to use non-linear disparity representation in the codec, so that each sample value is defined by the following power-law expression (1):

$$v = \left(\frac{1}{R} \cdot d\right)^{\alpha} \cdot R \qquad (1),$$

where v is non-linear representation (used in the codec) of normalized disparity value d, R is maximal normalized disparity value (e.g. 255 for 8-bit precision) and α is constant, experimentally found to be about 1.2÷1.4. In such way, closer objects are represented more accurately than the distant ones and thus quantization is non-uniform. Obviously, a reciprocal operation is performed after decoding the disparity map. As we will show later in the Experimental Results Section, this leads to improvement of the subjectively assessed quality of the reconstructed video.
It is interesting that that a research made in another context of synthesized video quality measured by PSNR led to somewhat different conclusions [7].

3. DISOCCLUSION REGION CODING

Often, the side views (both texture and depth) may be efficiently synthesized from the base central view. Usually there are only small portions of frames that cannot be synthesized as they are occluded in the central view (see Figs 1-2). Only these small regions that are disoccluded in the side view need to be coded while the remaining parts of the side view (both texture and depth map) may be accurately predicted from the central view.

In HEVC, pictures are coded in rectangular Coding Units (CUs). Therefore, in the side views, only the CUs containing disoccluded regions are coded. In our proposal, those regions are detected and used in side view encoders for compression. The shape of disoccluded regions is not directly transmitted - it is derived in the decoder from the reconstruction of central view depth map. Finally, at the decoder side, transmitted disoccluded regions are merged with views synthesized from the central view.
The transmission of disoccluded regions is done for both texture and depth of the side views. In our proposal it is performed by HEVC-based encoders with use of special CU-synth coding tool. In that tool, depending on the shape of disoccluded regions (and thus availability of synthesized content from the central view) some parts of the coded image are not coded. This is performed on CU-basis and thus shape of coded region must be aligned to CU boundaries (see Fig. 3). The largest CU blocks are composed from 64×64 pixels which can be split into smaller sub-CU blocks, according to quad-tree scheme, down to 4×4 pixels. In general, shape of disoccluded regions does not strictly fit into HEVC CU-blocks so principles for choosing proper CU size and split pattern has been developed.

In our proposal, basic principle is that CU containing no disoccluded pixels is forced not to be divided and no information (including split flag) is send in a bitstream to represent that CU. Such CU type is called CU-synth. Otherwise, if split of a CU block would result in three CU-synth blocks and one non-CU-synth block, split is forced but no split flag information is sent in a bitstream either, saving up bitrate. On the decoder side, this split flag can be derived in a similar way. In both described cases amount of information sent in a bitstream is considerably reduced. In other cases, if splitting of a CU would result in more than one non-CU-synth sub-CUs, HEVC rate-distortion-optimization mechanism makes decision on splitting or inter/intra coding the CU as is in generic HEVC.

Beside described advantages of the disoccluded regions coding it results in two sort of disadvantages. First is that arithmetic coder is likely to suffer from small number of context updating information when number of CU blocks are coded in CU-synth mode. Second is that CU blocks coded as inter or intra blocks (as in generic HEVC) that are surrounded by synthesized blocks have no motion prediction information. Despite this fact experimental results discussed in the Experimental Results Section prove that using of the disoccluded regions coding results bitrate reduction.

4. OTHER CODING TOOLS

The above described coding of disoccluded regions is a very efficient compression tool that brings significant reduction of the bitrate. Also nonlinear disparity representation allows remarkable reduction of bitrate with the guarantee of high subjective quality of the video. In the proposal [4] for 3D video compression technology, several additional compression tools have been used:

- Inter-view disparity compensation - a tool similar to one already used in MVC [8];
- Depth-Based Motion Prediction (DBMP) –motion vectors are predicted from the central view by a 3D mapping [9].
- Depth-dependent adjustment of QP for texture - QP parameter for texture is adjusted for CUs depending on the corresponding depth values.
- Depth-Gradient-based Loopback Filter (DGLF) and Availability Deblocking Loopback Filter (ADLF) - additional in-loop filters that reduce the artifacts resulted from coding of disoccluded regions.
- Residual layer coding – components relatedto high temporal frequency are separated from texture, and then they arejointly encoded for several views using few parameters related to the spectrum envelope.

The more detailed specification of the whole codec and the devised tools can be found in [4].

4. EXPERIMENTAL RESULTS

The proposed coding technology has been subjectively assessed by 13 test laboratories during resolution of MPEG group Call for Proposals (CfP) on 3D Video Coding Technology [3]. According to the results, our proposal was qualified as one of the best performing and showed a gain of about 3-5 MOS points relatively to the anchors, which were state-of-the-art HEVC coding technology (Fig. 5).

Unfortunately, these results do not bring any information about gains of the particular tools used in the proposal. Therefore, we have performed subjective evaluation in a way resembling MPEG tests that yielded in-depth view on performance of our tools coding [10].

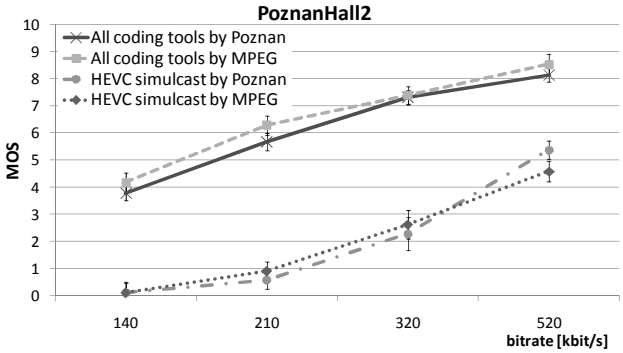

Figure 5. Comparison of subjective test results performed by MPEG and by Poznan University of Technology

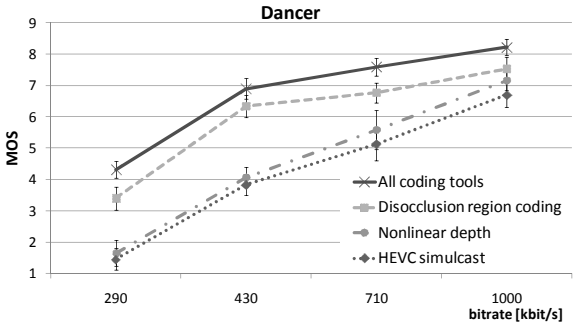

Figure 6. Subjective test results for Dancer sequence

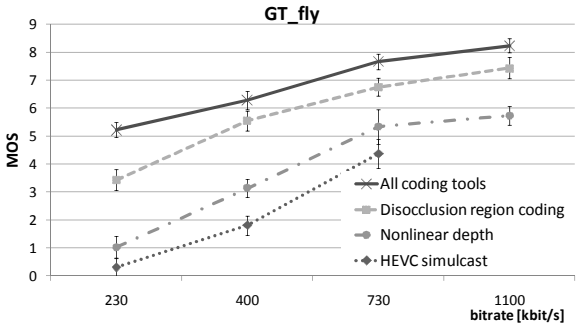

Figure 7. Subjective test results for GT_fly sequence

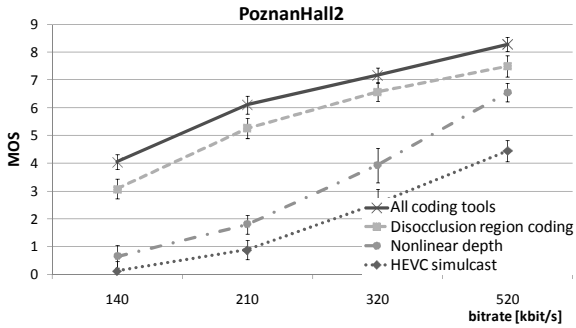

Figure 8. Subjective test results for PoznanHall2 sequence

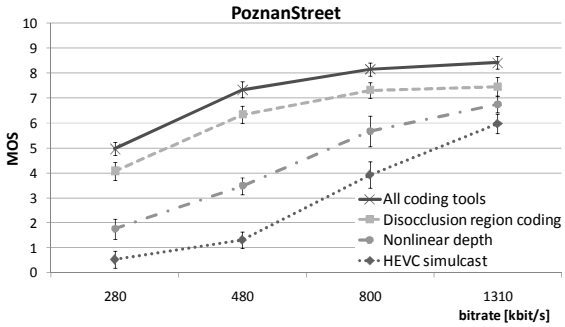

Figure 9. Subjective test results for PoznanStreet sequence

We have tested subjective quality of the proposed distinct tools (separately), performance of the full proposed codec and also MPEG anchors (HEVC simulcast). Figures 6-9 present results for 4 Full-HD sequences that show that most of the gain comes from disocclusion region coding tool (about 2-3 MOS points). Also, considerable gain can be attained with use on non-linear disparity representation (about 1-2 MOS points). The remaining tools provided a gain of less than 1 MOS point and were statistically undistinguishable.

Table 1. Objective results for central view.

Sequence	Improvement over anchor (Bjontegaard)		Amount of central view texture data in the whole 3D bitstream [%]
	ΔPSNR [dB]	ΔBitrate [%]	
PoznanHall2	2.5	-59.3	71,6
PoznanStreet	3.9	-69.6	64,0
Undo_Dancer	2.7	-62.6	71,4
GT_Fly	3.1	-67.0	71,3
Average of all 8 test sequences	**4.7**	**-69.8**	**62,8**

Also, we have analyzed our proposal objectively. Table 1 shows that average PSNR gain of 8 MPEG test sequences is about 4.7dB which corresponds to about 69% of bitrate saving relatively to HEVC-simulcast anchor. Another interesting observation is that the texture of the central view consumes about 60-70% of the whole bitstream (Table 1). That means that two side views (texture and depth) plus depth of the central view consume only about 30-40% of the whole 3D video bitstream. Because there are two views considered in the experiment, bitrate of a single side view is less that 20% of the whole 3D video bitstream.

5. CONCLUSIONS

The proposed coding technology allows reduction of required bitrate of about 50-70% comparing to HEVC simulcast. Such substantial gain is achieved mostly by use of two proposed tools: disoccluded region coding and non-linear disparity representation, which allow reduction of amount of data transmitted for depths and in the side views. The attained bandwidth can be utilized to benefit of the central view which has a major influence of the subjective quality of the whole 3D video.

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, 2010.

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

[4] M. Domański, *et al.*, "Technical Desciption of Poznan University of Technology proposal for Call on 3D Video Coding Technology", ISO/IEC JTC1/SC29/WG11 (MPEG) M22697, Geneva, Switzerland, November 2011.

[5] 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.

[6] 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.

[7] T. Senoh, K. Yamamoto, R. Oi, Y. Ichihashi, T. Kurita, „Proposal on non-linear normalization of Depth maps to 8 bits”, ISO/IEC MPEG m21189, Torino, Italy, 2011.

[8] J. Stankowski, *et al.*, "Extensions of HEVC technology for efficient multiview video coding", ICIP 2012, *to be published.*

[9] J. Konieczny, M. Domański, “Extended Inter-View Direct Mode for Multiview Video Coding”, IEEE Int. Conf. Acoustics Speech SignalProc., 2011, Prague, Czech Republic, May 2011.

[10] F. Lewandowski, M. Paluszkiewicz, „Quality subjective assessment for stereoscopic video”, BSc Thesis, Poznan University of Technology, Poland, January 2012, *in polish.*