# NONLINEAR DEPTH REPRESENTATION FOR 3D VIDEO CODING

*Olgierd Stankiewicz, Krzysztof Wegner, Marek Domański*

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

## ABSTRACT

The paper deals with nonlinear transformations of depth-sample values and their applications to multiview-plus-depth video coding. It is proposed to apply such a transformation prior to depth coding. For this transformation, the experiments demonstrate improved quality of the virtual views that are synthesized with use of the decoded depth. Such experiments were made for two versions of the prospective extensions of the Advanced Video Coding (MPEG-4 part 10, ITU-T Rec. H.264). standard. The experimental results are given for the scenario where multiview video is encoded independently form depth and also for the scenario where multiview video is coded jointly with depth.

***Index Terms***— 3D video, depth coding, depth representation, multiview plus depth representation.

## 1. INTRODUCTION

Nowadays, the 3D video systems are evolving from simple stereoscopic systems to sophisticated second-generation systems that provide more realistic perception of the 3D space [1]. Prospective applications of the second-generation 3D video systems include autostereoscopic displays, variable-baseline-distance systems as well as the free-viewpoint television [2,3]. The second-generation 3D video systems need efficient representations of 3D scenes. Practical description of a 3D scene is multiview video plus depth (MVD) [4], i.e. multi-viewpoint video together with the corresponding depth maps estimated during the process of content production.

Therefore, the compression of depth was identified as an important research task. This task is somewhat different from the task of video compression where the goal is to compress visual data in such a way that the decoded video is possibly similar to the input video. On the contrary, depth is not watched by a viewer but it is used to synthesize virtual views needed for an autostereoscopic display or in a free-viewpoint television system. So, mostly the decoded depth quality is expressed by the quality of the synthesized views.

In the standardization expert groups like MPEG, VCEG and JCT-3V, there is strong expectation to use the existing coding tools as much as possible for coding of depth. The basic compression tools are usually capable of processing 8-bit samples, and they use uniform quantization, i.e. the quantization with the constant quantization step that might be changed for some data structures like slices and macroblocks or coding units. Such uniform quantization is characteristic for both basic modern video standards: the AVC [5] and the new one – the HEVC [6]. On the other hand, it can be intuitively understood that exact depth is very important for foreground objects while small depth degradations in far background are mostly well tolerated by the human visual system. Therefore, nonuniform quantization would probably be appropriate for depth coding. Therefore, in order to preserve conformance with the standards like AVC and HEVC, we propose to process the depth values using a nonlinear function. Such processing together with uniform quantization is equivalent to the requested nonuniform quantization.

The idea of depth processing using nonlinear transformation of the depth-sample values is not a new one. It was already considered in [7] but with no particular relation to compression. In [8], a nonlinear transformation of sample values was used to obtain finer depth quantization in the background, i.e. a nonlinear transformation was used in the opposite way to that proposed in this paper.

The authors of this paper have already used the nonlinear depth representation in their 3D video compression technology proposed for MPEG in the response to Call for Proposals for 3D Video Coding Technology. This proposal was very successful as it was scored as one of the best performing codecs among those proposed to MPEG. This coding technology has been already described elsewhere [9-12] while here we will focus on the nonlinear depth representation itself.

The organization of this paper is the following. In Section 2, the general idea of nonlinear depth representation is considered while coding applications of nonlinear depth representation are considered in Section 3. Practical implementation of nonlinear depth representation is described in Sections 4 and 5 in the same way as it was used for standardization in video coding. The increases of the coding performance coming from the nonlinear depth representation are demonstrated in Section 6.

## 2. THE IDEA OF NONLINEAR DEPTH REPRESENTATION

Assume that distance to a point on a real object is $z$. Practical limitations yield that for all objects in a scene the depth values are within a final interval, i.e. $z_{min} < z < z_{max}$ , where $z_{min}$, $z_{max}$ are the distances to closest and the farthest object in a scene. The depth data are usually stored as inverted data $D$ [4]

$$D = round\left[255 \cdot \left(\frac{1}{z} - \frac{1}{z_{max}}\right) \Big/ \left(\frac{1}{z_{min}} - \frac{1}{z_{max}}\right)\right]. \qquad (1)$$

For uniform quantization, already this representation has the following advantage: a higher depth resolution of nearby objects is obtained. In order increase this effect, an additional nonlinear transformation is proposed to be performed on the depth-sample values

$$E = F[D], \qquad (2)$$

where $E$ is the transformed depth and $F$[.] is a nonlinear function, e.g. as shown in Fig.1 for the most common case of 8-bit representations.

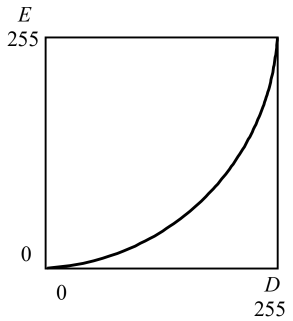


Fig.1. Nonlinear depth transformation performed before coding.

The transformation (2) is performed on depth samples before coding. Now, the depth coding itself is performed on the internal values $E$ instead of the external values $D$. This nonlinear transformation influences prediction errors and their linear transforms (mostly DCT-like) that are used in the course of the intraframe and interframe coding. The transform samples are quantized and this process is influenced by the proposed nonlinear depth transformation.

In our response to Call for Proposals for 3D Video Coding Technology issued by MPEG, we have successfully used the nonlinear transformation defined by Eq. (3)

$$E = \left(\frac{D}{D_{max}}\right)^{a} \cdot E_{max} \quad , \qquad (3)$$

where $D_{max}$ and $E_{max}$ are the maximum values of D and E, respectively. It was shown experimentally that already simple choice $a = 1.3$ (for small QP values) and $a = 1.6$ (for large QP values) gives good results. The choice of the nonlinear transformation (2) will discussed more in detail in Section 5.

After the depth decoding the depth samples must be processed by the inverse nonlinear transformation

$$E = F^{-1}[D] . \qquad (4)$$

Application of the nonlinear transformations should be optional. Therefore the following information should be signaled to the decoder:

- nonlinear transformation flag (signaling if the transformation is used),
- nonlinear transformation definition (sent if the abovementioned flag is set on).

The amount of additional information is quite small as it is proposed to transmit it per sequence.

The transformations (2) and (4) are done through LUT tables, and in some cases (view synthesis etc.) the respective LUT table may be incorporated into other LUT tables that must be used anyway. Therefore, the complexity of the nonlinear depth transformations is negligible.

## 3. NONLINEAR DEPTH REPRESENTATION IN 3D VIDEO CODING

Currently, multiview plus depth coding is subject of intensive standardization works. In the course of standard preparations by the expert groups MPEG and JCT-3V, two major scenarios of depth coding have been formulated:

1. Depth is compressed independently from multiview video in that sense that depth does not influence coding and decoding of multiview video. Such an approach is used in the depth coding extensions of the multiview coding techniques: MVC [13] and MVHEVC (MHVC) [14]. For MVC, this approach is supported by ATM-HP test model software.
2. Depth values are used in the course of video coding and decoding, e.g. for view-synthesis prediction. Such an approach is used in the 3D video coding proposed as extensions of the AVC [15] and HEVC [16] standards. For AVC, this approach is supported by ATM-EHP test model software.

In the first approach (Fig. 2), the transformations (2) and (4) do not influence any encoding or decoding process. Therefore the information about nonlinear depth transformation may be transmitted in the SEI (supplementary enhancement information) messages. The depth coding extension of MVC [13] has already incorporated *depth representation information SEI message* that optionally may be used to transmit the information about depth transformation.

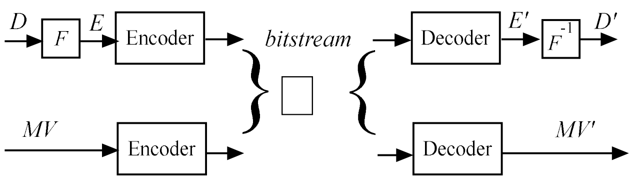


Fig.2. Independent depth coding: $E'$ and $D'$ denote decoded values, and $MV$ and $MV'$ are original and decoded multiview video, respectively.

In the second approach (Fig. 3), the encoding and decoding of multiview video exploits the information about depth. A good example of such a depth-dependent operation is view-synthesis prediction. Such prediction needs the values of *D* rather than *E*. Therefore the values of the external representation *D* must be used in the course of multiview video encoding and decoding (see Fig. 3).

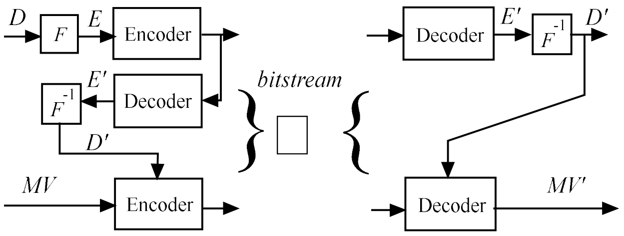


Fig. 3. Depth-dependent coding.

## 4. STANDARD DEFINITION OF THE NONLINEAR TRANSFORMATION

The shape and the parameters of the transformation (2) may be optimized individually by the encoder. Also further developments may bring new ideas about the definitions of the transformation. Therefore standards have to use flexible but simple means to define the transformation. Therefore it was decided (e.g. [13]) that the function will be linearly approximated in the intervals. It was accepted that a set of the equidistant deviations from the diagonal (see Fig. 4) [17]. The advantage of this type of definition is that the approximation can be easily calculated using fixed-point arithmetic.

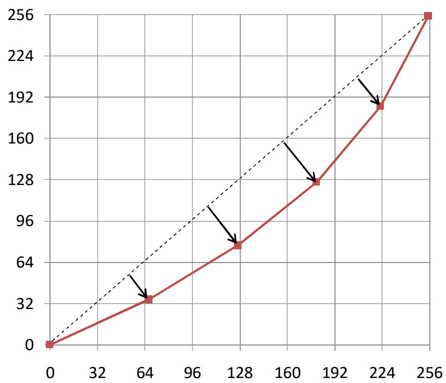


Fig. 4. Transformation definition by equidistant deviations and the linear approximation in the intervals

## 5. PRACTICAL IMPLEMENTATION OF THE NONLINEAR DEPTH TRANSFORMATION

The assumption is that the quantization step *q* decreases with increasing of depth *D*. For small values of *D* (far objects) the quantization step *q* is assumed to be large, while for large values of *D* (close objects) the quantization is expected to be fine. If ($D_{max}$-$D_{min}$) is normalized to one, there is

$$q(D) = Ae^{-\alpha D}, \quad (5)$$

where α is a parameter (e.g. α = 100 is a typical choice used in the experiments by the authors). As the sum of all the quantization steps should cover the whole unit interval

$$1 = \int_0^1 q(\delta)\, d\delta = A \int_0^1 e^{-\alpha\delta} d\delta = -\frac{A}{\alpha}(e^{-\alpha} - 1) \;, \quad (6)$$

where $\delta$ is the integration variable used instead of *E* or *D*, and the parameter *A*

$$A = -\frac{\alpha}{e^{-\alpha} - 1} \;. \quad (7)$$

Thus, the inverse nonlinear depth transformation

$$D = F^{-1}(E) = \int_0^E q(\delta) d\delta = -\frac{A}{\alpha}(e^{-\alpha E} - 1) \;, \quad (8)$$

and after some mathematical operations we get the forward transformation

$$E = F(D) = -\frac{1}{\alpha}\ln(1 - D(1 - e^{-\alpha})) \;. \quad (9)$$

The transformations (8) and (9) have been used in the experiments described in the next section.

Moreover, if the distribution of depth *D* samples is concentrated around small part values of *D*, usually it is better to switch off the nonlinear depth transformation. Such abnormal depth distributions are identified when center *C* of mass of *D* distribution is below a predefined threshold that was set to 100 in the experiments with 8-bit samples. This relatively simple condition can be used for automatic switching the tool on and off for individual sequences.

## 6. EXPERIMENTAL RESULTS FOR AVC AND MVC EXTENSIONS

The goal of the experiments was to estimate gains resulting from nonlinear depth representation. The experiments have been performed for some sequences that have passed the abovementioned for distribution of *D*.

The experiments have been done for depth coding extensions of the AVC standard [5]. These extensions have been already mentioned in Section 3: this is ATM-HP codec that works according to the first scenario, i.e. with independent depth coding [13] and ATM-EHP codec that exploits depth for more efficient encoding of multiview video [15].

In all experiments, we have used the transformation defined in Section 5 with δ = 100. The transformation was implemented by approximation with 41 nodes, i.e. deviations have been defined for 39 nodes (for two boundary nodes the deviation is always 0, for *D*=0 and *D*=255). The deviation vector (see Fig. 4) is the following:
2;4;7;8;10;12;14;16;17;19;20;21;22;23;24;25;26;26;27;27;27;27;27;27;26;26;25;24;23;22;20;19;17;15;13;11;9;6;3.

Table 1. Bjøntegaard gains in bitrate and PSNR due to application of nonlinear depth representation.

| Sequence | Multiview video coding | | Depth coding | | 3 views with depth maps | | 3 views with depth maps and 6 synthesized views | |
|---|---|---|---|---|---|---|---|---|
| | dBR [%] | dPSNR [dB] | dBR [%] | dPSNR [dB] | dBR [%] | dPSNR [dB] | dBR [%] | dPSNR [dB] |
| ATM-HP | | | | | | | | |
| GT Fly | 0.00 | 0.00 | -21.93 | 1.37 | -1.25 | 0.05 | -0.11 | 0.00 |
| Kendo | 0.00 | 0.00 | -25.76 | 1.64 | -5.28 | 0.28 | -4.13 | 0.20 |
| Balloons | -0.01 | 0.00 | -25.87 | 1.34 | -3.27 | 0.17 | -2.59 | 0.13 |
| ATM-EHP | | | | | | | | |
| GT Fly | 1.36 | -0.05 | -23.87 | 1.49 | 0.63 | -0.02 | 0.74 | -0.02 |
| Kendo | 0.19 | -0.01 | -20.35 | 1.30 | -3.41 | 0.15 | -3.13 | 0.14 |
| Balloons | 0.44 | -0.02 | -21.79 | 1.13 | -1.35 | 0.06 | -1.18 | 0.05 |

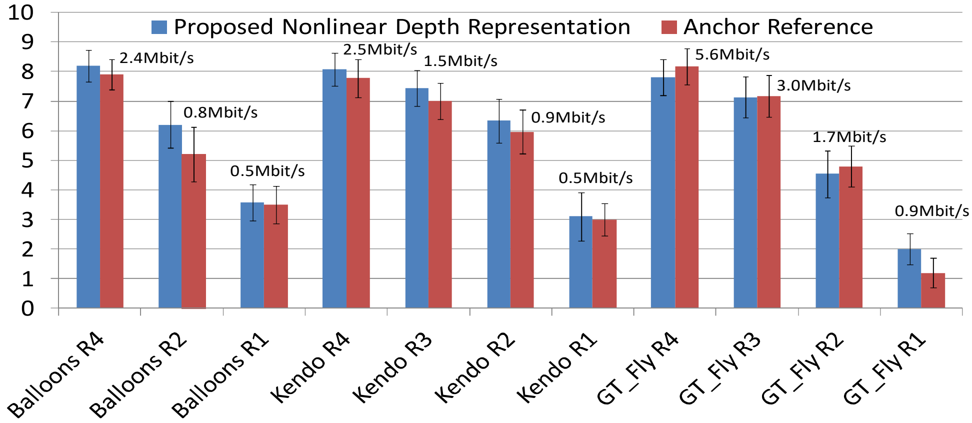


Fig. 5. Subjective quality assessment for the synthesized video encoded without ("anchor") and with ("proposed NDR") nonlinear depth representation. R1 to R4 denote different bitrates according to MPEG test conditions [20].

### 6.1. Measurements using objective video quality

The results are summarized in Table 1. The average gains in bitrate (dBR) as well as in PSNR (dPSNR) were calculated using Bjøntegaard measures [19]. The gains were calculated for:

- Video-only bitrate (depth bitstream not included) and average PSNR for 3 encoded views;
- Depth-only bitrate and PSNR for depth (average for depth maps for 3 views);
- Total bitrate for 3 views and 3 depth maps and average PSNR for 3 decoded views;
- Total bitrate for 3 views and 3 depth maps and average PSNR for 3 decoded views and 6 synthesized views.

The bitrate reduction is up to 4% while there is no measurable increase of complexity.

Moreover, the tests have been done also for the test sequences that do not fulfill the requirements for the depth distribution as described in Section 5 (for those sequences the transformation was switched off). The clear result of this experiment is that nonlinear depth representation does not decrease coding efficiency in such cases.

### 6.2. Measurements using subjective video quality

Also subjective test have been performed in order to compare visual quality of the synthesized views produced from the compressed depth maps both in the presence and in the absence ("anchor reference") of nonlinear depth representation (for the same bitrate). For the tests, 32 subjects have assessed the quality of stereo clips (2 subjects needed to be rejected) using the single stimulus method. The results for various bitrates (R1 – R4) are depicted in Fig. 5 together with 95% confidence intervals. The bitrates have been selected according to MPEG guidelines for individual test sequences [20]. The results show that nonlinear depth transformation improves coding efficiency.

## 7. CONCLUSIONS AND FURTHER DEVELOPMENTS

In this paper we have discussed various issues related to nonlinear depth representation. We have also demonstrated the benefits yielded by application of such representation.

Nonlinear depth representation is considered for 3D video coding standardization. The idea has been also adopted by other research groups who are developing it. First of all, adaptive slice-based switching of nonlinear depth transformation has been proposed in [21].

**Acknowledgement**
This work was supported by the public funds by National Science Centre, Poland, as a research project DEC-2012/05/B/ST7/01279.

# REFERENCES


[1] D. Wood, "The truth about stereoscopic television", *International Broadcasting Convention IBC 2009*, Conference Publication, Amsterdam, 2009.

[2] M. Tanimoto, M. P. Tehrani, T. Fujii, T. Yendo, "Free-viewpoint TV", *IEEE Signal Processing Magazine*, vol. 28, January 2011, pp. 67-76.

[3] M. Tanimoto, "FTV: Free-viewpoint Television", Signal Processing: Image Communication, vol. 27, Issue 6, July 2012, pp. 555–570.

[4] K. Müller, P. Merkle, T. Wiegand, "3-D video representation using depth maps", *Proceedings of the IEEE*, vol. 99, no. 4, April 2011, pp. 643-656.

[5] ISO/IEC 14496-10, Int. Standard, "Generic coding of audio-visual objects – Part 10: Advanced Video Coding," 7th Ed., 2012, [also:] ITU-T Rec. H.264, Edition 7.0, 2012.

[6] B. Bross, W.-J. Han, J.-R. Ohm, G. Sullivan, Th. Wiegand, "Study Text of ISO/IEC DIS 23008-2 High Efficiency Video Coding" *ISO/IEC JTC1/SC29/WG11*, Doc. N13155, Shanghai, Oct. 2012.

[7] I. Feldmann, O. Schreer, P. Kauff, "Nonlinear depth scaling for video applications", *Proceedings of 4th Workshop on Digital Media Processing for Multimedia Interactive Services*, E. Izquierdo (ed.), World Scientific, 2003, pp. 433-438.

[8] T. Senoh, K. Yamamoto, R. Oi, Y. Ichihashi, T. Kurita, "Proposal on non-linear normalization of depth maps to 8 bits", *ISO/IEC JTC1/SC29/WG11,* MPEG Doc. M21189, Torino, Italy, July 2011.

[9] M. Domański, O. Stankiewicz, K. Wegner, et al. "Technical Description of Poznan University of Technology proposal for Call on 3D Video Coding Technology", *ISO/IEC JTC1/SC29/WG11*, Doc. M22697, Geneva, CH, Nov. 2011.

[10] M. Domański, O. Stankiewicz, K. Wegner et al. "New Coding Technology for 3D Video within Depth Maps as Proposed for Standardization within MPEG", *Int. Conf. Systems, Signals and Image Processing IWSSIP 2012*, Vienna, April, 2012, pp. 415-418.

[11] M. Domański, O. Stankiewicz, K. Wegner et al. "Coding of Multiple Video+Depth Using HEVC Technology and Reduced Representations of Side Views and Depth Maps", *29th Picture Coding Symposium PCS 2012*, Kraków, May 2012, pp. 5-8.

[12] M. Domański, O. Stankiewicz, K. Wegner et al. "3D video compression by coding of disoccluded regions", 2012 *IEEE Int. Conf. Image Proc. ICIP 2012*, Orlando, Sept. 2012.

[13] T. Suzuki, M. M. Hannuksela, Y. Chen, G. Sullivan, S. Hattori, "Study Text of ISO/IEC 14496-10:2012/DAM2 MVC extension for inclusion of depth maps", *ISO/IEC JTC1/SC29/WG11*, Doc. N13140, Shanghai, Oct. 2012.

[14] G. Tech, K. Wegner, Y. Chen, M. Hannuksela, " Multi-view HEVC Working Draft Text 2", *ISO/IEC JTC1/SC29/WG11*, Doc. N13159, Shanghai, Oct. 2012.

[15] M. Hannuksela, Y. Chen, T. Suzuki, J.-R. Ohm, G. Sullivan, "Text of ISO/IEC 14496-10:2012/PDAM3 AVC compatible video-plus-depth extension", *ISO/IEC JTC1/SC29/WG11*, Doc.N13142, Shanghai, Oct. 2012.

[16] G. Tech, K. Wegner, Y. Chen, S. Yea, "3D-HEVC Test Model 2", *ISO/IEC JTC1/SC29/WG11*, Doc. N13160, Shanghai, Oct. 2012.

[17] O. Stankiewicz, K. Wegner, J. Siast, "3D-CE2a results on nonlinear depth representation", *ISO/IEC JTC1/SC29/WG11*, Doc. M25017, Geneva, May 2012.

[18] O. Stankiewicz, K. Wegner, "Results of subjective evaluation in 3DV-CE2", *ISO/IEC JTC1/SC29/WG11*, Doc. M25244, Geneva, May 2012.

[19] Bjøntegaard G.: *Calculation of average PSNR differences between RD-curves*, ITU-T Study Group 16 Question 6 (VCEG), Doc. VCEG-M33, Austin TX, March 2001.

[20] "Common Test Conditions for AVC and HEVC-based 3DV" *ISO/IEC JTC1/SC29/WG11*, Doc. N12560, San Jose, USA, February 2012.

[21] I. Lim, H. Wey, D. Park, "3D-CE7.a Improved Nonlinear Depth Representation", *Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Doc. JCT3V-C0094, Geneva, Jan. 2013.