IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 14, 2019, accepted February 28, 2019, date of publication March 5, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2903087

Study of 3D Video Compression Using Nonlinear Depth Representation

OLGIERD STANKIEWICZ, (Member, IEEE), KRZYSZTOF WEGNER, (Member, IEEE), AND MAREK DOMAŃSKI, (Member, IEEE)
Chair of Multimedia Telecommunications and Microelectronics, Poznań University of Technology, 60-965 Poznań, Poland
Corresponding author: Olgierd Stankiewicz (olgierd.stankiewicz@put.poznan.pl)

This work was supported by the National Science Centre, Poland, under Grant DEC-2012/05/B/ST7/01279.

ABSTRACT The extensions of Advanced Video Coding (AVC) and High-Efficiency Video Coding (HEVC) for multiview video plus depth allow for the coding of 3D video scenes. During the standardization of these extensions, it was demonstrated that compression performance can often be improved by the application of nonlinear depth transformation prior to the compression and applying the respective backward nonlinear depth transformation after decompression. Such an approach is referred to as the compression of nonlinear depth representation (NDR). In this paper, a survey of the NDR technology is provided, but a large part of this survey comprises yet unpublished results. First, the motivation and rationale behind the idea of NDR are discussed. The options for the choice of nonlinear transformation are provided with the respective formulations. Also, the standardization of 3D video coding with NDR within ISO/IEC and ITU is discussed along with the experimental results that demonstrate benefits related to the application of NDR in 3D-AVC and 3D-HEVC. Furthermore, the experimental comparisons are provided for simulcast coding of multiview plus depth video using AVC and HEVC technology with and without the NDR. The similar experimental results are also provided for simulcast coding using the current version of Versatile Video Coding technology that is under development for the forthcoming standard of ISO/IEC (MPEG-I) and ITU. All of the experimental results have been obtained for the standard JVC-3V video test sequences under common test conditions for 3D video coding.

INDEX TERMS 3D video, 3D-HEVC, AVC, coding, compression, depth coding, depth representation, MVD representation, Versatile Video Codec, VVC.

I. INTRODUCTION

3D video compression is strongly related to the compression of depth maps, which should be considered in the context of their exploitation in the system. Depth maps may be used directly in order to calculate the distance to obstacles, e.g., in navigation systems of mobile robots [1], or may be used in the synthesis of virtual views, i.e. in Depth-Image-Based Rendering (DIBR) [2], [3]. The latter is substantial for emerging communication applications where compression is crucial. These applications include Free-viewpoint TeleVision (FTV) and virtual navigation in a scene [3], [4], as well as 3D video gaming and content delivery to glasses-free 3D displays, such as autostereoscopic, super-multiview, lightfield or holographic displays (e.g. [5]), or immersive media [6]. For these and similar application scenarios, the coding errors in depth maps should influence the quality of the synthesized virtual views as weakly as possible. Therefore, in this paper, the quality of the virtual views is used as a metric in order to assess the compression performance for the depth maps, as in [7]–[10].

For 3D video, the most widely used video format is Multiview Video plus Depth (MVD) [11]. The video and depth components may be coded independently (simulcast coding) or jointly. For the latter case, there exist two options:

1. The depth maps are coded first and are then used for the view-synthesis-based prediction of the remaining views [9], [10], [12].
2. The views are coded before the depth maps, and the coding of depth maps exploits information from the views [9], [13]–[15]. This approach was standardized by ISO and ITU as 3D HEVC [16], [17] that is the current state-of-the-art compression technology for MVD.

The associate editor coordinating the review of this manuscript and approving it for publication was Madhu S. Nair.

For MVD video coding, the recent investigations concern optimal mode selection [18], as well as optimal bit allocation between views and depth maps [19], [20].

As the depth is more vulnerable to degradation of the edges than is the natural image or video, depth-specific coding methods were developed. The state-of-the-art techniques for depth coding include platelets [21]–[23] and wedgelets [24], [25]. The general idea of platelets is that inside a given coded region of depth (e.g., corresponding to a macroblock or a coding unit), the depth is modeled as a flat plane called a platelet [21]–[23]. In [26], the platelets are considered in the context of AVC-based (Advanced Video Coding) multi-view coding. Wedgelets extend the platelets in such a way that a given block of depth samples is represented using multiple planes separated by edges [27]. The applications of platelets and wedgelets in depth coding are considered in [24], [25], and [28]. Other depth coding techniques are synthesis-based [8], [13], [14], based on compressed sensing [29], or else exploit predictions between MVD components [30]–[32].

Unfortunately, for several of the aforementioned techniques of depth compression (e.g., those based on platelets or wedgelets) the implementations are different from those for general-purpose monoscopic codecs. This is significant especially at the low level of the video coding layer, where the processing speed must be extremely high. From the implementation point of view, high-level depth coding tools that exploit low-level structures from general-purpose codecs are highly desirable. Indeed, depth consumes about 10% or less of the total 3D video bitrate [9], [10], thus the fingerprint of depth-specific tools should be minimal.

The research on MVD coding techniques has also been stimulated by the standardization of ISO/IEC and ITU, which coordinated the activities of the respective expert groups: MPEG (Motion Pictures Experts Group) and VCEG (Video Coding Experts Group) and worked together in a Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V). In 2011, in response to a Call for Proposals (CfP) on 3D Video Coding Technology [33], several interesting codecs were proposed as the starting point for two tracks of standardization, aimed at extending Advanced Video Coding (AVC) [34] and High Efficiency Video Coding (HEVC) [16] monoscopic video coding standards, respectively. In 2015, the AVC standard has been amended by extension of its Multiview Video Coding part by the depth coding tools (so called MVC+D) as well by new 3-D video coding technology (3D-AVC), currently used in 3-D High profile [12]. Moreover, the HEVC standard has been amended by the MV-HEVC technology that does not have explicit tools for depth coding [17]. The encoding of depth in MVD representation was addressed with by the 3D-HEVC which also has been included in HEVC standard.

One of the tools studied during the standardization processes for AVC and HEVC is Nonlinear Depth Representation (NDR) [35]. The idea of NDR is to transform the depth values before processing and compression, and then to restore the original depth values before virtual view synthesis, i.e. to compress the transformed samples of depth instead of the depth itself (depth in this paper is synonymous to normalized disparity). Therefore, NDR involves coding a signal that is varied in a nonlinearly with respect to depth, so that, for instance, close objects are given a higher dynamic range than farther ones. The idea of NDR is close to that of video contrast correction, also known as gamma correction, which is widely used in video processing and compression, e.g., [37].

NDR is assumed to provide the two fundamental features:

1. The forward and backward transformations related to NDR are simple and their implementations have very low complexity.
2. For most of the mentioned MVD compression methods including trivial application of monoscopic compression methods, compression of NDR is, on average more efficient than compression of direct depth representations.

The idea of nonlinear representation of depth has already been proposed by Krishnamyrthy *et al.* in 2001, but in their paper [38] it is mentioned rather briefly, e.g., no methodology for the selection of parameters of the nonlinear transform function of normalized disparity is given. Also, the paper considers only still image coding with JPEG2000. Therefore, at that time, the idea had no impact on 3D video coding standards.

In the context of aforementioned video coding standardization, the use of NDR has been proposed by the authors of this paper already in their coding technology proposal [10], [36] for the MPEG's "Call for proposals on 3D video coding technology" [33]. Later, during a series of so-called core experiments, NDR has been tested and finally accepted as a part of the extensions of the AVC standard: for depth compression (so called MVC+D), and 3D-AVC [34].

This paper is devoted to study of NDR and mainly its implementations in current 3D video standards. Nevertheless, the paper goes beyond the results that are published in previous works, in which study on NDR is rather brief [10], does not include the rationale behind the idea of NDR, is related to AVC video coding technology only [35], or is dissolved in documents in MPEG and JCT-3V databases [36], [47], [48], [52], [53], [64], [65]. In the present work NDR is studied holistically, in the context of the most important state-of-the-art video and 3D video coding technologies, including Versatile Video Coding (VVC) [39]–[41] - a new video coding technology that is currently being prepared and is planned to be standardized during until year 2020. In the present paper, the rationale behind the idea of NDR is considered, and the derivation of the transform functions is studied in detail, as this is an issue not included into standards.

The main contributions of this paper include:

- Presenting a holistic, previously unpublished study of NDR and survey of its applications within the current technologies: MVC+D, AVC-3D, HEVC, HEVC-3D and VVC.

- Providing novel results for compression with NDR on top of HEVC-3D and VVC video coding technologies.
- Providing exhaustive results, of both subjective and objective evaluation, and discussion of the results among different codecs.

II. NONLINEAR TRANSFORMATION OF DEPTH

The straightforward approach to depth map representation is to use uniformly quantized disparity values, linearly normalized to range $0 \ldots 2^N - 1$ (e.g., $0 \ldots 255$ for N = 8 bits per sample), commonly called depth. Unfortunately, for a finite number of bits per sample, a uniform quantization is often accompanied by annoying artifacts in the synthesized views. A uniform quantization of disparity does not match the properties of the human visual system that is more tolerant to disparity errors in the background of a synthesized scene (Fig. 1) than to errors in the foreground.

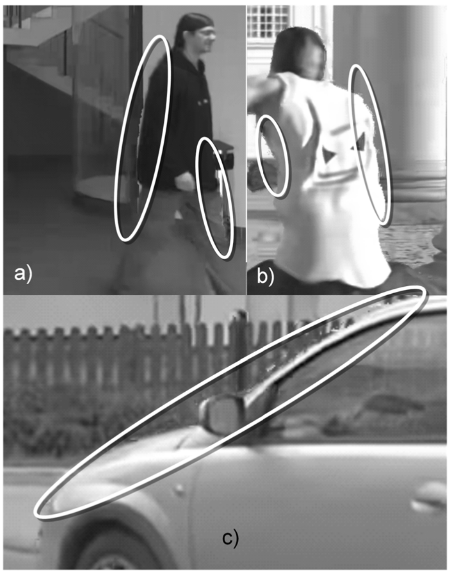

FIGURE 1. Examples of artifacts related to linear quantization in coding of depth maps, especially visible for objects in the foreground, marked with ellipses. a) "Poznan Hall 2" sequence, b) "Undo Dancer" sequence, and c) "Poznan Street" sequence.

A. CONCEPT OF NONUNIFORM QUANTIZATION

In terms of the motivating rationale of the present study, it is beneficial to develop a scheme that resembles nonuniform quantization such that the depth values for close objects, (especially the foreground objects) are given a finer quantization than the depth for distant objects. A direct implementation of such a nonuniform quantization of depth would be inconvenient in the current video coding technologies such as AVC or HEVC, where the residual signal is quantized rather than the original depth signal. The residual signal results from complex processes involving transformation (e.g. Discrete Cosine Transform), prediction (intra or inter), in-loop filtering and rate-distortion optimization (also influencing quantization). In video codecs, residual signal quantization procedures are related to well-established and highly developed procedures for bitrate and quality control. Therefore, the modification of the disparity or depth quantization process should not influence the inner quantization in the video codecs. Therefore, the modification of the disparity or depth quantization process should not influence the inner quantization in the video codecs.

Also, in standardization expert groups such as MPEG, VCEG, and JCT-3V, there was a strong expectation [42] to use existing monoscopic video coding tools as much as possible for the coding of depth. The basic compression tools that are featured in both modern video standards (AVC [34] and HEVC [16]) are capable of processing N-bit unsigned samples and employ uniform (or semi-uniform) quantization, which is not optimal for coding depth that is expected to be used for virtual view synthesis. Therefore, in order to preserve conformance with standards such as AVC or HEVC, in the formulation of the NDR technique it was proposed to process depth values using a nonlinear function. Such processing together with uniform quantization is equivalent to the requested nonuniform quantization (Fig. 2).

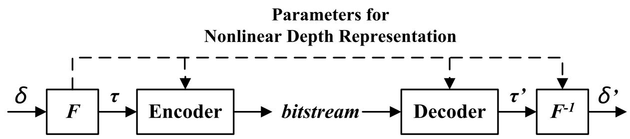

FIGURE 2. Nonuniform depth quantization realized using the nonlinear transform F performed on the input depth representation δ and the inverse nonlinear transform F^{-1} applied to the output of the codec.

B. ASSUMPTIONS AND DEFINITIONS

Let us assume that the distance from the camera along its optical axis 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 finite range, i.e., $z_{near} < z < z_{far}$, where z_{near} and z_{far} are the distances to the nearest and the farthest objects in a scene. Depth data is usually stored as normalized disparity δ which is inversely proportional to the depth [43]:

$$\delta = \delta_{max} \cdot \left(\frac{1}{z} - \frac{1}{z_{far}}\right) / \left(\frac{1}{z_{near}} - \frac{1}{z_{far}}\right), \tag{1}$$

where $\delta_{max} = 2^N - 1$ is the maximal value of δ represented by N-bit samples. Although N > 10 is desirable for many depth-image based rendering (DIBR) applications [44], still N = 8 is often used.

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

$$\tau = F(\delta), \tag{2}$$

where τ is the transformed depth and $F(\cdot)$ is a nonlinear function, e.g., as shown in Fig. 3 for the most common case of 8-bit samples.

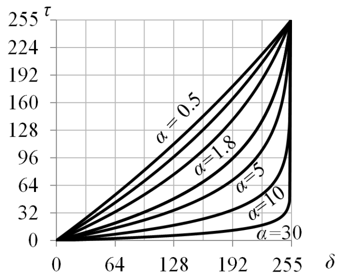

FIGURE 3. Family of the nonlinear depth transformation functions (17) for $\alpha \in \{0.5, 1.0, 1.8, 3.5, 5.0, 10.0, 30.0\}$.

The transformation (2) is performed on depth samples before coding. Now, the depth coding itself is performed on the internal values τ instead of the external values δ. This nonlinear depth transformation has an effect on prediction errors and their linear transforms (mostly DCT-like) that are used in the course of intra-frame and inter-frame coding. The transform samples are quantized and this process is influenced by the proposed nonlinear depth transformation.

After transmission and decoding, the inverse nonlinear transformation $F^{-1}(\cdot)$ is applied to the decoded transformed depth τ', and the reconstructed normalized disparity δ' is retrieved:

$$\delta' = F^{-1}(\tau'). \tag{3}$$

In the next subsection, suitable possibilities for the nonlinear transformation F will be presented.

C. POWER-LAW FUNCTION

At first, the authors of the present paper have proposed a shape form for the nonlinear transformation function [10] that was inspired by the idea of gamma-correction that is well-known from classic video technology [37]:

$$\tau = \left(\frac{\delta}{\delta_{max}}\right)^{\gamma} \cdot \tau_{max}, \tag{4}$$

where δ_{max} and τ_{max} are the maximal values of δ and τ, respectively (e.g., 255 for 8-bit samples).

In [10], it was shown experimentally that the compression performance can be increases by a simple choice of γ in the range 1.2–1.6, depending on the quantization step ($\gamma \approx 1.6$ for large quantization steps, $\gamma \approx 1.2$ for very fine quantization).

The aforementioned proposal for nonlinear depth transformation (4) has been implemented as a part of the 3D video codec [10] developed by the Poznań University of Technology, Chair of Multimedia Telecommunications and Microelectronics in response to the "Call for Proposals on 3D Video Coding Technology" issued by the MPEG group in 2011 [33]. This proposal has been rated very high among other proposals and was found to be one of the best performing proposals in the HEVC category. The excellent results attained by the proposed codec provoked a deeper analysis of the shares of the gains provided by the particular proposed coding tools [45], [46]. At the time of initial proposal, Equation (4) was used without a theoretical background; in the current paper a novel approach with a theoretical derivation has been undertaken, as presented below in detail.

D. EXPONENTIAL FUNCTION

Here, the theoretical derivation of the nonlinear depth transformation $F(\delta)$ is provided. For the sake of simplicity, instead of considering the values of δ and τ (within the ranges $0 - \delta_{max}$, and $0 - \tau_{max}$, respectively) and the transform function F, let us first consider $\tilde{\delta}$ and $\tilde{\tau}$ which are normalized to the interval [0, 1]:

$$\tilde{\delta} = \frac{\delta}{\delta_{max}}; \quad \tilde{\tau} = \frac{\tau}{\tau_{max}}. \tag{5}$$

The forward and inverse nonlinear depth transformations can be defined for $\tilde{\delta}$ and $\tilde{\tau}$ similarly as in (2) and (3):

$$\tilde{\tau} = \tilde{F}\left(\tilde{\delta}\right); \quad \tilde{\delta} = \tilde{F}^{-1}(\tilde{\tau}). \tag{6}$$

The first requirement (R1) is that distant objects are quantized less finely than the closer ones, as intended from the nonlinearity of the $\tilde{F}$ transform. Let us denote it as quantization step function $s\left(\tilde{\delta}\right)$ such that:

$$\tilde{\delta} = \tilde{F}^{-1}(\tilde{\tau}) = \int_0^{\tilde{\tau}} s(u)du. \tag{7}$$

Notably, due to a different selection of z_{near} and z_{far} values, the position of scene objects within the depth/disparity range can vary. Therefore, function $s\left(\tilde{\delta}\right)$ should not depend on depth position but rather on the size of the object.

Basing on the abovementioned, we define the second requirement (R2) that $s\left(\tilde{\delta}\right)$ is position-invariant, such that a constant shift $\Delta\tilde{\delta}$ of disparity $\tilde{\delta}$ results in the same relative change of quantization step $s\left(\tilde{\delta}\right)$, independently from the value of $\tilde{\delta}$, depending on $\Delta\tilde{\delta}$ only:

$$\frac{s\left(\tilde{\delta} + \Delta\tilde{\delta}\right)}{s\left(\tilde{\delta}\right)} = f(\Delta\tilde{\delta}). \tag{8}$$

where f is a some single-argument function of $\Delta\tilde{\delta}$ only.

Let us consider the following exponential function that (as we will show later) conforms to both (R1) and (R2) requirements mentioned above:

$$s\left(\tilde{\delta}\right) = A \cdot e^{-\alpha \cdot \tilde{\delta}}. \tag{9}$$

$\alpha > 0$ is a constant parameter, for which range $\alpha \in [0.5, 5.0]$ is a good choice for experiments; for $\alpha \leq 0.5$ the nonlinear depth transformation function becomes practically linear and for $\alpha > 5.0$ it becomes impractically bended (see Fig. 3).

The value of parameter A can be easily derived from (7), knowing that both the domain and codomain of $\tilde{F}$ and $\tilde{F}^{-1}$

are interval [0, 1]:

$$1 = \int_0^1 s(u)du = A \cdot \int_0^1 e^{-\alpha \cdot u} du = -\frac{A}{\alpha} \cdot \left(e^{-\alpha} - 1\right), \quad (10)$$

where u is the integration variable used in place of $\tilde{\delta}$ or $\tilde{\tau}$. Therefore, A can be expressed as follows:

$$A = \frac{\alpha}{1 - e^{-\alpha}}. \quad (11)$$

Notably, for the assumed $\alpha > 0$, the value of parameter A is greater than 1.

Now, as the values of α and A are defined, we can analyze the proposed exponential quantization step function $s\left(\tilde{\delta}\right)$ (9) regarding the requirements (R1) and (R2) mentioned above:

R1) Distant objects are quantized more coarsely than closer ones; for small values of $\tilde{\delta}$ (far objects) the quantization step $s\left(\tilde{\delta}\right)$ is large, while for large values of $\tilde{\delta}$ (close objects) the quantization is fine. Thus, the requirement (R1) is met.

R2) The proposed $s\left(\tilde{\delta}\right)$ is position-invariant; (12) follows directly from (8) and (9). Therefore, requirement (R2) is also met:

$$\frac{s\left(\tilde{\delta} + \Delta\tilde{\delta}\right)}{s\left(\tilde{\delta}\right)} = \frac{A \cdot e^{-\alpha \cdot \left(\tilde{\delta} + \Delta\tilde{\delta}\right)}}{A \cdot e^{-\alpha \cdot \tilde{\delta}}} = e^{-\alpha \cdot \Delta\tilde{\delta}} = f(\Delta\tilde{\delta}). \quad (12)$$

Knowing that the requirements are met, we can derive the inverse nonlinear depth transformation $\tilde{F}^{-1}$ as:

$$\tilde{\delta} = \tilde{F}^{-1}(\tilde{\tau}) = \int_0^{\tilde{\tau}} s(u)du = A \cdot \int_0^{\tilde{\tau}} e^{-\alpha \cdot k} du, \quad (13)$$

$$\tilde{\delta} = -\frac{A}{\alpha} \cdot \left(e^{-\alpha \cdot \tilde{\tau}} - 1\right). \quad (14)$$

After some mathematical operations we can get $\tilde{\tau}$:

$$\tilde{\tau} = -\frac{1}{\alpha} \cdot ln\left(-\frac{\alpha}{A} \cdot \tilde{\delta} + 1\right), \quad (15)$$

which yields the forward nonlinear depth transformation $\tilde{F}$. After substitution of A and some further simplifications:

$$\tilde{\tau} = \tilde{F}\left(\tilde{\delta}\right) = -\frac{1}{\alpha} \cdot \ln\left(1 - \tilde{\delta} \cdot \left(1 - e^{-\alpha}\right)\right). \quad (16)$$

Finally, in the original notation (without scaling to unit interval (6)) we obtain the desired forward transformation

$$\tau = F(\delta) = -\frac{\tau_{max}}{\alpha} \cdot \ln\left(1 - \frac{\delta}{\delta_{max}} \cdot \left(1 - e^{-\alpha}\right)\right). \quad (17)$$

The results attained with the use of transformation (17) lead not only to comparable subjective gains as in the case of power-law-based expression (4), but also provide objective gains as measured using Peak Signal-to-Noise Ratio (PSNR). This fact has been brought to the attention of the MPEG group [47] for consideration as a tool for a new generation of coding technology standards.

III. STANDARDIZATION OF NDR

In this section, the process of adopting of Nonlinear Depth Representation into international video coding standards will be discussed. In particular, approximation of the transform function F for the sake of compression will be considered along with application schemes of NDR into different codecs.

A. SEGMENT-WISE APPROXIMATION OF THE TRANSFORM FUNCTION F

The best practice is for standards to leave as much flexibility as possible to their implementers. Therefore, in a video coding standard, it would not be wise to define a single arbitrary transformation, e.g., defined by (4), or (17). In order to fulfill the desire for flexibility, the MPEG group has accepted the authors' proposal [34] that the definition of function $F(\cdot)$ is provided to the encoder, then transmitted in the bitstream and finally used in the decoder. It was decided to define $F(\cdot)$ by its approximation using a polygonal chain (polygonal line). Therefore, in a standard bitstream, only equidistant deviations from the diagonal $\tau = \delta$ are transmitted, i.e. the deviation vector: $\boldsymbol{w} = [w_0, w_1, \ldots, w_N]$ is transmitted (see Fig. 4) [47].

The flexible definition of the nonlinear depth transformation exhibits several advantages:

- The definition is generic and it generalizes all the aforementioned nonlinear transformations.
- The transformation may be optimized in the encoder according to the depth content.
- Further developments may result in nonlinear depth transformations that also can be defined as in Fig. 4.
- The proposed approximation can be easily calculated using fixed-point arithmetic [34].

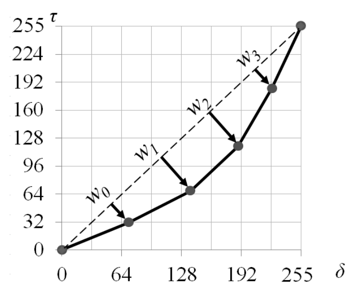

FIGURE 4. Nonlinear function definition with the use of the equidistant deviations w_i and the linear approximation in the intervals (as used in AVC extensions [34]). In this example, the deviation vector $\boldsymbol{w}$ is a 4-element vector [35].

B. CONTROL MECHANISM FOR NONLINEAR DEPTH REPRESENTATION

The experiments performed during the standardization procedure indicated that for some content types the nonlinear depth transformation brings no gain, and it is useful to disable this tool for that content. In particular, it was noticed (see Section IV) that, if the distribution of the normalized disparity δ is concentrated in only a small part of interval, the $[0, \delta_{max}]$, it is usually better to switch off the nonlinear depth transformation. Such abnormal depth distributions can be identified on the basis of the expected value $\mathrm{E}\left[\delta_{\mathrm{x,y}}\right]$ of the normalized disparity $\delta_{x,y}$:

$$E\left[\delta_{x,y}\right]=\frac{1}{W\cdot H}\sum_{x=1}^{W}\sum_{y=1}^{H}\delta_{x,y}, \tag{18}$$

where W and H correspond to the width and height of the image, respectively, and x, y are the coordinates of the depth samples in the image (a video frame).

Therefore, it is beneficial to enable a nonlinear depth transformation only when $\mathrm{E}\left[\delta_{x,y}\right]$ is above a predefined threshold value:

$$Enable\ NDR=\begin{cases}\text{false} & \text{if } \mathrm{E}\left[\delta_{x,y}\right]<E_{threshold}\\ \text{true} & \text{if } \mathrm{E}\left[\delta_{x,y}\right]\geq E_{threshold},\end{cases} \tag{19}$$

where $E_{threshold}$ has been experimentally set to 100 for the case of 8-bit samples of the depth, independently from compression technology used: AVC-, HEVC- or VVC-based. This relatively simple condition can be used for automatically switching the tool on and off for individual sequences based upon the decisions made for some key frames.

The mentioned advantage of flexibility of nonlinear depth representation proposal can be used in order to improve compression efficiency, e.g. by adaptive selection of parameters [48].

C. APPLICATION OF NDR IN CODECS

In the course of compression standard preparations by MPEG and JCT-3V expert groups, two major scenarios of depth coding have been considered:

1. Depth is compressed independently from multiview video such that depth does not influence the coding and decoding of multiview video.
2. Depth values are used in the course of video coding and decoding.

In the first approach (Fig. 5a), the nonlinear depth transformation does not influence the encoding or decoding of views. Therefore, the information about the nonlinear depth transformation may be transmitted in Supplementary Enhancement Information (SEI) messages. Therefore, the depth coding extension of MVC called MVC+D [34] has already incorporated NDR information in an SEI message that may be used to transmit the information about the optional depth transformation.

In the second approach (Fig. 5b), the encoding and decoding of multiview video exploits information about depth. A good example of such a depth-dependent operation is view-synthesis prediction [49]. Performing such a prediction requires the values of normalized disparity rather than the transformed representation. Therefore the δ' values should be internally calculated using (3) both in the view encoder and decoder. In this case, the signaling of the nonlinear depth transformation must be included in mandatory part of bitstream (in contrast to SEI messages which need not be

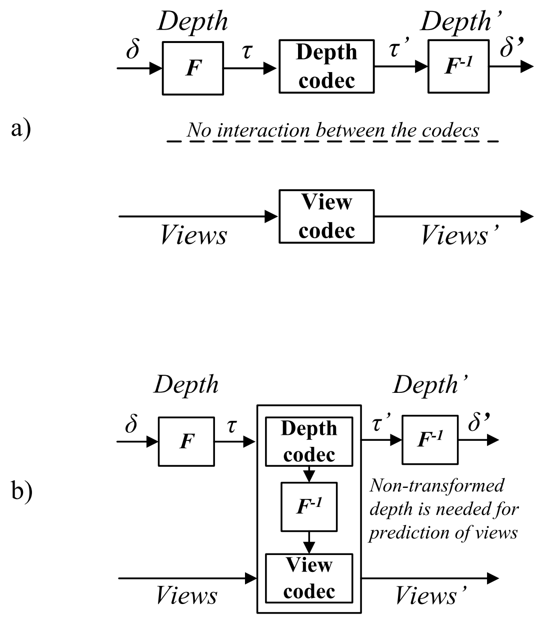

FIGURE 5. a) Independent depth and view coding and b) Depth-dependent coding of views. δ and τ denote the original normalized disparity map and transformed, nonlinearly represented values, δ' and τ' are decoded (reconstructed) values and *Views* and *Views'* are original and decoded multiview video, respectively.

implemented in the decoder). Such an approach is used in 3D video coding adopted to the AVC standard (3D-AVC [34]) and the HEVC standard (3D-HEVC [16]).

IV. EXPERIMENTAL ASSESSMENT

The goal of the experiments is to estimate the coding gains owing to application of the nonlinear depth transformation for various coding scenarios for multiview-plus-depth (MVD) video. These scenarios are individually defined for MVD compression using AVC, HEVC, and the current version of the forthcoming VVC technology.

A. METHODOLOGY OF EXPERIMENTS

For AVC- and HEVC-based compression, the nonlinear depth representation tool (under the name NDR) has been evaluated by MPEG and JCT-3V (experts in a series of experiments [50]–[53], executed according to the methodology defined in the Common Test Conditions (CTC) [7] document, and previously used for the evaluation of responses to the Call for Proposals on 3D Video Coding. These experiments were extended by the authors using the same methodology. Here, we summarize the results obtained by the authors partially in cooperation with MPEG and JCT-3V groups.

In the experiments, seven MPEG/JCT-3V test MVD sequences (Table 1) are used, all with 8-bit depth samples. For each test sequence, 3 views (“1”, ”2” and “3” in Fig. 6) are compressed along with the respective depth maps for four distinct bitrates (rate points R1,. . ., R4). Using the decoded views and the decoded depth maps, six virtual views (the views “v1”,. . .,”v6” in Fig. 6) are synthesized with the use of

TABLE 1. MVD test sequences used for experiments.

Sequence name		Type	Resolution	Provider
S1	Poznan Hall2	Natural/indoor	1920×1080	Poznań University of Technology [55]
S2	Poznan Street	Natural/outdoor	1920×1080	
S3	Undo Dancer	Synthetic/indoor	1920×1080	Nokia [56], [57]
S4	GT Fly	Synthetic/outdoor	1920×1080	
S5	Kendo	Natural/indoor	1024×768	Nagoya University [58]
S6	Balloons	Natural/indoor	1024×768	
S7	Newspaper	Natural/indoor	1024×768	GIST [59]

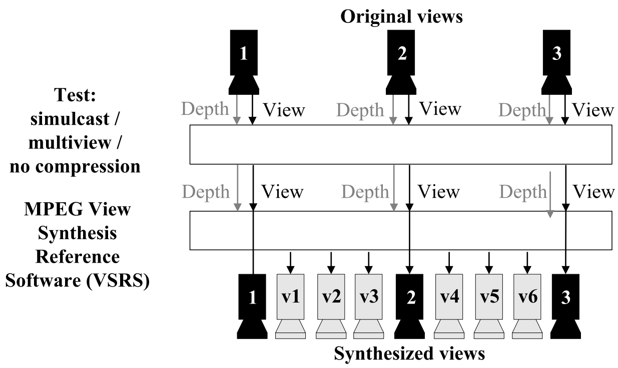

FIGURE 6. The arrangement of views: the real views are marked in black, while the views synthesized in the receiver ("v1"…"v6") are marked in gray.

TABLE 2. Bitrates [kbps] used in video coding experiments.

	AVC				HEVC				VVC			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
S1	750	900	1300	2300	210	310	480	770	70	120	230	490
S2	750	1100	1800	4000	410	710	1180	1950	140	230	420	840
S3	1380	1750	2300	2900	430	780	1200	2010	220	400	780	1700
S4	2000	2380	2900	4000	340	600	1080	1600	160	290	560	1200
S5	800	1000	1300	1900	280	430	670	1040	140	220	380	710
S6	500	600	800	1250	300	480	770	1200	140	230	380	700
S7	500	700	1000	1350	340	450	680	900	140	230	400	760

MPEG View Synthesis Reference Software (VSRS) [54]. The synthesis is also performed from the original, uncompressed views, as a reference for comparison.

The bitrate is always the total bitrate for 3 encoded views and 3 encoded depth maps. The rate points R1, …, R4 used in the experiments are selected individually for each test sequence and for each coding technology, i.e. AVC, HEVC, and VVC, in order to allow reliable subjective evaluation, e.g., the encoded video has quite low quality, not suitable for commercial broadcast, but allowing observation of both degradation and enhancement. Obviously, the bitrates selected for HEVC are lower than those for AVC, and those for VVC ([39]–[41]) are even lower (Table 2).

The results are given as the bitrate reductions averaged over all frame in a sequence, while sustaining the same average quality (luma PSNR or Mean Opinion Score: MOS). The gains are calculated using Bjøntegaard measures [60] between the codecs with an NDR tool and the results of the original, unmodified codecs without that tool.

The luma PSNR is calculated with respect to the views synthesized from the uncompressed views versus the views synthesized from the uncompressed depth maps. The luma PSNR averages are calculated over all six synthesized views ("v1", …,"v6" in Fig. 6) and over all frames from each view.

The subjective tests have been performed on a polarization monitor (Hyundai S465D) in accordance with the general rules of ITU Recommendation BT.500 [61]. Double Stimulus Impairment Scale (DSIS) methodology was used and the viewers were viewing interchangeably the reference (stereo pair synthesized from uncompressed original) and the tested codec (stereo pair synthesized from the compressed views and depths). The stereo pair was composed from views "v3" and "v4" (Fig. 6).

In all experiments the nonlinear depth transformation defined in (17) has been used with $\alpha = 1.8$. The transformation was defined by approximation with 41 nodes and therefore the deviations have been defined for 39 nodes (for the two boundary nodes with normalized disparity values $\delta = 0$ and $\delta = 255$, the deviation is trivially 0). The deviation vector w (see Fig. 4) that has been used is as follows:

$$w = [2, 4, 7, 8, 10, 12, 14, 16, 17, 19, 20, 21, 22, 23, 26, 27, 27, 27, 27, 27, 27, 26, 26, 25, 24, 23, 22, 20, 19, 17, 15, 13, 11, 9, 6, 3]. \quad (20)$$

B. NDR IN 3D-EXTENSIONS OF AVC

The first group of experiments in which NDR was evaluated during standardization, covered so-called HP and EHP profiles of AVC-based 3D video coding, which later became MVC+D and 3D-AVC, respectively. For the sake of brevity we will describe them here using the latter, more widely known names. The results have been submitted to MPEG in [62] and independently cross-checked [63]. The tests were performed with the use of JCT-3V 3DV-ATM v8.0 software. In Table 3 is presented a summary of the attained bitrate reductions. It can be seen that in some of the sequences (Balloons, Kendo, Newspaper) considerable gains are observed, in some (GT Fly, Poznan Street) small gains or losses are observed and in others (Poznan Hall 2, Undo Dancer) losses are observed.

The observation summarized in Table 3 inspired our work on a scheme that would adaptively switch the NDR tool on or off. The result of the research is described in Section IV B. The use of formula (19) resulted in turning on NDR in the case of three of the sequences from the set (GT Fly, Kendo and Balloons), while for others (Poznan Street, Poznan Hall 2, Undo Dancer, Newspaper) it has been turned off. The value of $E_{threshold}$ was set to 100. The adequately averaged results are also shown in Table 3.

The attained PSNR-based bitrate reduction reaches 4.13% (for Kendo sequence) while there is no measurable increase of complexity. On average over all of the test sequences (including sequences that do not fulfill the requirements for

TABLE 3. Bjøntegaard bitrate reductions (positive values), attained owing to the application of nonlinear depth representation in MVC+D and 3D-AVC.

Sequence name	Bitrate reduction (const. quality)			
	PSNR		MOS	
	MVC+D	3D-AVC	MVC+D	3D-AVC
1. GT Fly	0.11%	- 0.74%	0.54%	1.08%
2. Balloons	2.59%	1.18%	18.23%	14.05%
3. Kendo	4.13%	3.13%	5.56%	6.14%
Average: sequences 1-3	**2.28%**	**1.19%**	**8.11%**	**7.09%**
4. Poznan Street	0.65%	- 0.44%	0%	0%
5. Poznan Hall 2	- 4.43%	- 5.67%	0%	0%
6. Undo Dancer	- 3.23%	- 1.93%	0%	0%
7. Newspaper	1.54%	1.04%	0%	0%
Average: all 7 sequences	**0.19%**	**0.49%**	**3.48%**	**3.04%**
Average*	**0.98%**	**0.51%**	**3.48%**	**3.04%**

* Averaged results for Sequences 1–3 with NDR turned on, and of sequences 4–7 with zero gains owing to (19) and NDR turned off. In the case of MOS results, the viewing was not performed [62] and thus the MOS gains are trivially zeroes.

depth distribution, as described in Section III B, for which the transformation was switched off) the gains are 0.98% (MVC+D) and 0.51% (3D-AVC). When considered are only those sequences in which the transform has been turned on, the gains are 2.28% (MVC+D) and 1.19% (3D-AVC) on the average.

Note that the mentioned gains are coming solely from coding tools for depth, while depth constitutes about 10% of the whole bitstream.

As a part of evaluation of NDR during the MPEG meeting in Geneva in May 2012, subjective tests [62] have also been performed by experts. These test were conducted in order to compare the visual quality of the synthesized views produced from the compressed depth maps both in the presence and in the absence of NDR, both for the same bitrate. The tested codecs were:

- 3D-AVC anchor, or
- 3D-AVC with proposed NDR.

The tested sequence was always coded at a constant bitrate: from the highest (R4) to the lowest (R1), reflecting the Common Test Condition (CTC) [7] and the general methodology developed by MPEG for exploration experiments (EE) [64]. The bitrates (Table 2) have been selected according to MPEG guidelines for individual test sequences [7], [33].

The results for various bitrates (R1…R4) are depicted in Fig. 7, together with 95% confidence intervals. The results show that nonlinear depth transformation improves coding efficiency, although some of the confidence intervals overlap. The evaluation was performed only for the sequences for which NDR was turned on: GT Fly, Balloons, and Kendo (Fig. 7). Thus, the gains with respect to MOS constant quality for other sequences has not been measured (Table 3).

C. NDR USED WITH HEVC AND 3D-HEVC

In the context of HEVC-based 3D video coding, we have considered two base codecs:

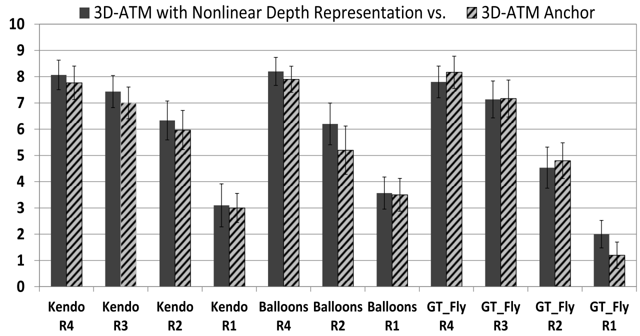

FIGURE 7. Results of subjective evaluation of nonlinear depth representation in the context of 3D-AVC video compression.

TABLE 4. Bjøntegaard bitrate reductions (reductions are positive values), attained owing to the application of nonlinear depth representation in HEVC (simulcast) and in 3D-HEVC.

Sequence name	Bitrate reduction (const. quality)			
	PSNR		MOS	
	HEVC	3D-HEVC	HEVC	3D-HEVC
1. GT Fly	21.25%	0.24%	27.78%	1.52%
2. Balloons	13.35%	1.29%	33.27%	15.39%
3. Kendo	13.07%	2.17%	43.91%	22.50%
Average: sequences 1-3	**15.89%**	**1.23%**	**34.99%**	**13.14%**
4. Poznan Street	21.40%	0.38%	35.85%	11.54%
5. Poznan Hall 2	13.59%	- 19.76%	23.37%	3.57%
6. Undo Dancer	19.94%	- 5.58%	7.68%	- 1.96%
7. Newspaper	9.59%	1.84%	17.40%	10.26%
Average: all 7 sequences	**16.03%**	**- 2.77%**	**27.04%**	**8.97%**
Average*	**6.81%**	**0.53%**	**14.99%**	**5.63%**

* Averaged results for Sequences 1–3 with NDR turned on, and of sequences 4–7 with zero gains due to (19) and NDR turned off.

- HEVC, using JCT-VC reference software HM v16.18),
- 3D-HEVC, using JCT-3V HEVC-3D reference software HTM v16.3, based on HM v16.18.

In both of those we have tested if the usage of NDR can bring coding gains. The tested coding scenarios were:

- HEVC simulcast,
- HEVC simulcast with Nonlinear Depth Representation,
- 3D-HEVC,
- 3D-HEVC with Nonlinear Depth Representation.

In our experiments, the number of subjects involved was greater than in the official MPEG evaluation. The relatively large number of subjects resulted in 95% confidence intervals were in the range of ±(0.05–0.15) and thus very small. Therefore, those intervals were not depicted on the plots (Fig. 8). It can be seen (Table 4) that the bitrate reductions in the case of HEVC-simulcast are higher than in the case of AVC-based coding. They are in the range of about 9%–21% (about 16% on average) if the same PSNR quality is considered, and in the range of about 7%–44% if the same MOS quality is considered. Thus, the new results are comparable with our previous research [10], even though a different version of the HEVC codec and nonlinear transform has been used.

In the case of 3D-HEVC the results are definitely worse than in the case of HEVC. If the same PSNR quality is considered, even some significant losses are observed

(Poznan Hall 2: about 20% increase of bitrate), which on average brings a loss of about 2%. If constant MOS quality is considered, NDR performs a little better and the gains are of about 3%–22% of bitrate reduction, 9% on average. Notably, the newly attained HEVC-3D results are also worse that those reported in our previous research for 3D video coding proposal [10] based on HEVC, presumably due to presence of more advanced depth-related compression tools present in 3D-HEVC.

We have also considered the scheme in which NDR is adaptively switched on and off, according to formula (19). The adequately averaged results are also presented in Table 4. It can be seen that for the worst-case sequences NDR has been disabled, which results in better average performance.

For those sequences for which NDR remained switched on (GT Fly, Kendo, Balloons), the average gains in 3D-HEVC are about 1% (PSNR) and 14% (MOS). For the whole set of sequences (including those sequences for which NDR is disabled and thus there were no gains nor losses) the bitrate reductions for 3D-HEVC are about 0.5% (PSNR) and about 5% (MOS).

The fact that the bitrate reductions for 3D-HEVC are smaller than for HEVC simulcast, may be result of additional depth coding tools in 3D-HEVC that already provide some gain in depth substreams.

D. NDR WITH VERSATILE VIDEO CODEC

We have also tested a technology that is currently being developed within the MPEG group in a project aiming for a new video coding standard under the name of Versatile Video Codec (VVC) [39]–[41]. Currently, VVC is not considered for 3D content in the MVD format but for monoscopic video only. However, it can be used to approximate the gains that can be attained with NDR, just as HEVC can be used to estimate the gains of 3D-HEVC. Therefore, the experiments that we have performed resembled those for HEVC.

The views and depth have been coded as separate streams, with the use of VVC reference software VVC Test Model 2 (VTM2) configured according to the JVET Common Test Conditions [65] in the following tested variants:

- VVC simulcast without NDR.
- VVC simulcast with NDR.

In Table 5 and Fig. 9 presented are the attained results. If MOS subjective quality is considered, the usage of NDR provides noticeable gains for most of the sequences. On average over all sequences, the bitrate reduction while preserving the same subjective quality is about 8%. On the other hand, if objective quality is considered, the usage of NDR brings a loss of about 1.4% on average over all sequences. Similarly, like in HEVC, this mainly results from the influence of the worst-performing sequences. A scenario where those sequences are excluded because of adaptive switching of NDR on and off, according to formula (19), is summarized in Table 5. For those sequences for which NDR remained switched on (GT Fly, Kendo, Balloons), the average gains in VVC are about 1.4% (PSNR) and 16.7% (MOS). For the entire set of sequences (including those sequences for which NDR is automatically disabled) the bitrate reductions for VVC are about 0.5% (PSNR) and about 7% (MOS).

TABLE 5. Bjøntegaard bitrate reductions (reductions are positive values), attained owing to the application of nonlinear depth representation in VVC simulcast.

Sequence name	Bitrate reduction (const. quality)	
	PSNR	**MOS**
1. GT Fly	0.18%	4.01%
2. Balloons	1.19%	28.72%
3. Kendo	2.67%	17.38%
Average: sequences 1-3	**1.36%**	**16.70%**
4. Poznan Street	- 0.04%	0.70%
5. Poznan Hall 2	- 11.61%	- 4.63%
6. Undo Dancer	- 2.82%	0.19%
7. Newspaper	0.63%	10.02%
Average: all 7 sequences	**-1.40%**	**8.06%**
Average*	**0.58%**	**7.16%**

* Averaged results of Sequences 1–3 with NDR turned on, and of sequences 4–7 with zero gains due to (19) and NDR turned off.

E. SUMMARY OF THE RESULTS

In this subsection we summarize and discuss the overall results attained for the use of Nonlinear Depth Representation in the context of different video coding technologies: MVC+D, AVC-3D, HEVC, HEVC-3D, and VVC video codecs.

Comparison of the gathered results presented in Table 6 leads to the conclusion that the coding gains attained with NDR depend on the level of advancement of other depth coding tools present in given codec. Notably, the greatest gains are attained on the basis of HEVC codec, which is missing any depth/3D-related tools; thus the depth is encoded as a regular video. If we consider bitrate reduction when constant MOS quality is assumed, the second greatest gains are attained with VVC, which also does not include any depth/3D-related tools. VVC, however, includes much more sophisticated prediction tools which presumably reduces PSNR gains of NDR to the moderate value of 1.23%. Moreover, inclusion of 3D/depth tools in AVC-3D and HEVC-3D, with respect to the base coding technologies, MVC+D and HEVC respectively, notably result in decreased gains attained with NDR.

Therefore it can be concluded that Nonlinear Depth Representation is a tool for depth compression that is competitive to other depth/3D-related tools included in the new 3D-video coding technologies – AVC-3D and HEVC-3D. The moderate PSNR gains attained in those cases, e.g. 1.19% of bitrate reduction in the case of AVC-3D, or 1.23% for HEVC-3D, are entirely justified if computational costs are taken into consideration. In order to assess this effect, computational complexity of the respective components of the system has also been evaluated during the experiments.

Table 7 presents the execution times of the encoder, of the decoder, and of the nonlinear depth transformation (the same for forward and inverse transforms) required for

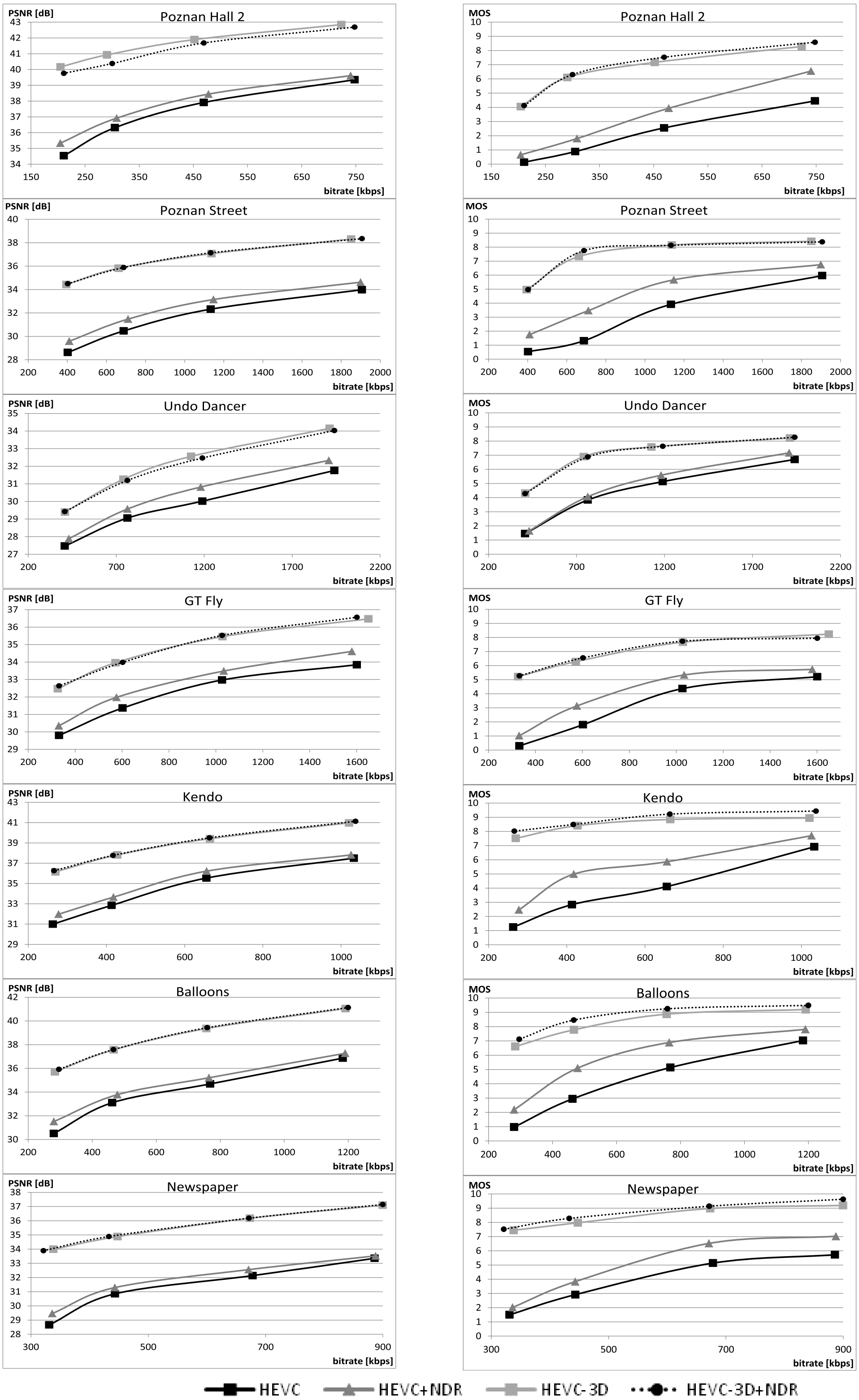

FIGURE 8. Quality comparison of HEVC-simulcast ("HEVC"), HEVC-simulcast with nonlinear depth representation ("HEVC+NDR"), 3D-HEVC, and 3D-HEVC with NDR ("3D-HEVC+NDR"). Left: Objective evaluation (PSNR). Right: Subjective evaluation (MOS). Confidence intervals of order of ±(0.05÷0.15) are omitted in the graphs the sake of clarity.

implementation of NDR. The results are presented per single frame, averaged over all test sequences. The test was performed using an Intel Core i7-3770K processor working at 3.5GHz. The execution-time share of NDR is negligible both with respect to the runtime of the encoder as well as with respect to the runtime of the decoder.

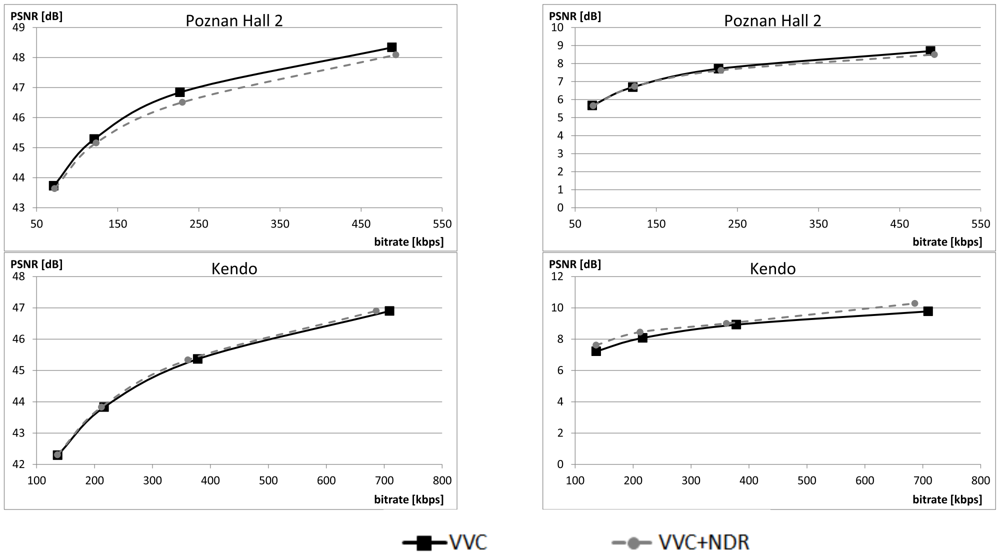

FIGURE 9. Quality comparison of VVC-simulcast ("VVC") against VVC-simulcast with nonlinear depth representation ("VVC+NDR"), Left: Objective evaluation (PSNR). Right: Subjective evaluation (MOS). Confidence intervals in the order of ±(0.08÷0.21) are omitted in the graphs for the sake of clarity. For the sake of brevity, only examples of the best and worst cases are shown.

TABLE 6. Summary of Bjøntegaard bitrate reductions (reductions are positive values), attained due to the application of nonlinear depth representation in different codecs. Results averaged over sequences s1-S3.

Codec	Bitrate reduction (const. quality)	
	PSNR	MOS
MVC+D	2.28%	8.11%
AVC-3D	1.19%	7.09%
HEVC	15.89%	34.99%
HEVC-3D	1.23%	13.14%
VVC	1.36%	16.70%

TABLE 7. Evaluation of computational complexity of nonlinear depth representation.

Process	Average execution time [ms]				
	MVC+D	AVC-3D	HEVC	HEVC-3D	VVC
Encoder	6093	8558	118810	62727	185140
Decoder	155	165	471	263	1120
NDR	5				

V. CONCLUSIONS

In this paper, an extensive survey on Nonlinear Depth Representation is provided. This representation of depth has been proposed by the paper authors in the context of a search for an efficient HEVC extension for Multiview plus depth video coding, known also as 3D video coding. Later, the paper authors have researched this representation during successive years.

The idea of NDR was already briefly mentioned in [10] together with some experimental results for HEVC. Then, some more results have been published in the context of AVC in the conference paper [35]. More results, mostly experimental results are dissolved in documents in MPEG and JCT-3V databases [36], [47], [48], [52], [53], [64], [65]. Many other results remained unpublished yet, as explained in Introduction.

The importance of NDR is related to its usage in the international standards on 3D video coding [16], [34]. In this paper, also prospective application of NDR to the forthcoming VVC compression technology is considered.

In this paper, for the first time, a generic Nonlinear Depth Representation is considered with respect to several aspects, including choice of the transformation and applications to several variants of the MVD video compression. The rationale for application of the nonlinear depth transformation is strongly related to the subjective quality of the synthesized views rendered using decoded view and depth maps. Therefore, it is not astonishing that the coding gains are larger when considering subjective quality measured as Mean Opinion Score (MOS) whereas the gains are moderate for quality measured by PSNR of the synthesized views.

For simulcast HEVC, the average bitrate reduction reaches even 27% for subjective quality assessment whereas it is only 16% for PSNR quality measurements. Even for simulcast coding using the current version of the novel highly efficient VVC technology, the average gain reaches 8% for subjective quality assessment. Smaller, but still mostly significant gains are achieved for special 3D and multiview+depth video coding techniques, both AVC- and HEVC-based.

The complexity of Nonlinear Depth Representation is very low. For the HEVC-based technology, on average it is much less than 1% of the total decoder complexity, and even less for encoders. Therefore, the relation between the complexity and bitrate reduction is very advantageous.

The average gains are reduced for some specific content for which we do not observe any significant bitrate reduction upon application of NDR. Fortunately, a simple control procedure was found, that allows for switching off the NDR tool when its application is inefficient. The decision is made upon simple analysis of a key frame.

The paper provides a general framework for nonlinear depth transformations that may be further developed. Moreover, the paper reviews the standardization of nonlinear depth representation. This issue was not considered in previous research papers.

Avenues for future research include development of an adaptive control mechanism that would allow for efficient use of nonlinear depth representation for a broader variety of content.

REFERENCES

[1] S. C. Diamantas, A. Oikonomidis, and R. M. Crowder, "Depth estimation for autonomous robot navigation: A comparative approach," in *Proc. IEEE Int. Conf. Imag. Syst. Techn.*, Thessaloniki, Greece, Jul. 2010, pp. 426–430.

[2] C. Fehn, R. Barre, and S. Pastoor, "Interactive 3-DTV-concepts and key technologies," *Proc. IEEE*, vol. 94, no. 3, pp. 524–538, Mar. 2006.

[3] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "FTV for 3-D spatial communication," *Proc. IEEE*, vol. 100, no. 4, pp. 905–917, Apr. 2012.

[4] O. Stankiewicz, M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch, and J. Samelak, "A free-viewpoint television system for horizontal virtual navigation," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2182–2195, Aug. 2018.

[5] Y. J. Jeong, J.-H. Lee, Y. H. Cho, D. Nam, D.-S. Park, and C. C. J. Kuo, "Efficient light-field rendering using depth maps for 100-mpixel multi-projection 3D display," *J. Display Technol.*, vol. 11, no. 10, pp. 792–799, Oct. 2015.

[6] M. Domański, O. Stankiewicz, K. Wegner, and T. Grajek, "Immersive visual media—MPEG-I: 360 video, virtual navigation and beyond," in *Proc. IEEE Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Poznań, Poland, May 2017, pp. 1–9.

[7] *Common Test Conditions for AVC and HEVC-Based 3DV*, document N12560, ISO/IEC JTC1/SC29/WG11, San Jose, CA, USA, Feb. 2012.

[8] J. Y. Lee and H. W. Park, "Efficient synthesis-based depth map coding in AVC-compatible 3D video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 6, pp. 1107–1116, Jun. 2016.

[9] K. Müller *et al.*, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3366–3378, Sep. 2013.

[10] M. Domański *et al.*, "High efficiency 3D video coding using new tools based on view synthesis," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3517–3527, Sep. 2013.

[11] K. Müller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proc. IEEE*, vol. 99, no. 4, pp. 643–656, Apr. 2011.

[12] M. M. Hannuksela *et al.*, "Multiview-video-plus-depth coding based on the advanced video coding standard," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3449–3458, Sep. 2013.

[13] H. Oh and Y.-S. Ho, "H.264-based depth map sequence coding using motion information of corresponding texture video," in *Advances in Image and Video Technology* (Lecture Notes Comput. Science), vol. 4319. Berlin, Germany: Springer-Verlag, Dec. 2006.

[14] B. T. Oh, J. Lee, and D.-S. Park, "Depth map coding based on synthesized view distortion function," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1344–1352, Nov. 2011.

[15] J. Y. Lee, H.-C. Wey, and D.-S. Park, "A fast and efficient multi-view depth image coding method based on temporal and inter-view correlations of texture images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 12, pp. 1859–1868, Dec. 2011.

[16] *High Efficiency Video Coding*, 3rd ed., Standard ISO/IEC IS 23008-2, Oct. 2017.

[17] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Jan. 2016.

[18] H. Yuan, S. Kwong, J. Liu, and J. Sun, "A novel distortion model and Lagrangian multiplier for depth maps coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 3, pp. 443–451, Mar. 2014.

[19] K. Klimaszewski, O. Stankiewicz, K. Wegner, and M. Domański, "Quantization optimization in multiview plus depth video coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 27–30.

[20] H. Yuan, Y. Chang, J. Huo, F. Yang, and Z. Lu, "Model-based joint bit allocation between texture videos and depth maps for 3-D video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 485–497, Apr. 2011.

[21] Y. Zhang, H. Wang, and J. Zhao, "Depth map compression based on platelet coding and quadratic curve fitting," in *Proc. IEEE 27th Can. Conf. Elect. Comput. Eng. (CCECE)*, Toronto, ON, Canada, May 2014, pp. 1–4.

[22] Y. Morvan, P. H. N. de With, and D. Farin, "Platelet-based coding of depth maps for the transmission of multiview images," *Proc. SPIE*, vol. 6055, Jan. 2006, Art. no. 60550K.

[23] Y. Morvan, D. Farin, and P. H. N. de With, "Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multiview images," in *Proc. IEEE Int. Conf. Image Process.*, San Antonio, TX, USA, Sep./Oct. 2007, pp. V-105–V-108.

[24] P. Merkle, C. Bartnik, K. Müller, D. Marpe, and T. Wiegand, "3D video: Depth coding based on inter-component prediction of block partitions," in *Proc. Picture Coding Symp. (PCS)*, Kraków, Poland, May 2012, pp. 149–152.

[25] K. Müller, P. Merkle, G. Tech, and T. Wiegand, "3D video coding with depth modeling modes and view synthesis optimization," in *Proc. Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Hollywood, CA, USA, Dec. 2012, pp. 1–4.

[26] P. Merkle *et al.*, "The effects of multiview depth video compression on multiview rendering," *Signal Process., Image Commun.*, vol. 24, nos. 1–2, pp. 73–88, Jan. 2009.

[27] M. Zhang, C. Zhao, J. Xu, and H. Bai, "A fast depth-map wedgelet partitioning scheme for intra prediction in 3D video coding," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Beijing, China, May 2013, pp. 2852–2855.

[28] F. Jäger, "Segment-wise prediction in 3D video coding," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2015, pp. 463–466.

[29] M. Sarkis and K. Diepold, "Depth map compression via compressed sensing," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 737–740.

[30] K.-J. Oh, J. Lee, and D.-S. Park, "High priority intra coding method for depth video coding," in *Proc. Picture Coding Symp.*, May 2012, pp. 45–48.

[31] K.-J. Oh, A. Vetro, and Y.-S. Ho, "Depth coding using a boundary reconstruction filter for 3-D video systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 3, pp. 350–359, Mar. 2011.

[32] S. Yea and A. Vetro, "Multi-layered coding of depth for virtual view synthesis," in *Proc. Picture Coding Symp. (PCS)*, Chicago, IL, USA, May 2009, pp. 1–4.

[33] *Call for Proposals on 3D Video Coding Technology*, Standard ISO/IEC JTC1/SC29/WG11 MPEG2011/N12036, Geneva, Switzerland, 2011.

[34] *Advanced Video Coding*, 8th ed., Standard ISO/IEC Int. 14496-10, Sep. 2014.

[35] O. Stankiewicz, K. Wegner, and M. Domański, "Nonlinear depth representation for 3D video coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Melbourne, VIC, Australia, Sep. 2013, pp. 1752–1756.

[36] M. Domański *et al.*, *Technical Desciption of Poznan University of Technology Proposal for Call on 3D Video Coding Technology*, document M22697, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Nov. 2011.

[37] C. Poynton, *Digital Video and HD*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 2012.

[38] R. Krishnamurthy, B.-B. Chai, H. Tao, and S. Sethuraman, "Compression and transmission of depth maps for image-based rendering," in *Proc. Int. Conf. Image Process.*, vol. 3, Thessaloniki, Greece, Oct. 2001, pp. 828–831.

[39] *Requirements for a Future Video Coding Standard v4*, document N16359, ISO/IEC JTC1/SC29/WG11 MPEG 115th MPEG Meeting, Geneva, Switzerland, Jun. 2016.

[40] V. Baroncini, P. Hanhart, M. Wien, J. Boyce, A. Segall, and T. Suzuki, *Results of the Joint Call for Evidence on Video Compression With Capability Beyond HEVC*, Standard ISO/IEC JTC1/SC29/WG11 N17054, Turin, Italy, Jul. 2017.

[41] J. Chen, Y. Ye, and S. H. Kim, *Algorithm Description for Versatile Video Coding and Test Model 2*, document ITU-T SG 16 WP 3 ISO/IEC JTC 1/SC 29/WG, vol. 11, JVET-K1002, Joint Video Experts Team, 11th Meeting, Ljubljana, Slovenia, Jul. 2018.

[42] *Applications and Requirements on 3D Video Coding*, document N11829, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Mar. 2011.

[43] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision*. New York, NY, USA: McGraw-Hill, 1995.

[44] J. Stankowski, Ł. Kowalski, J. Samelak, M. Domański, T. Grajek, and K. Wegner, "3D-HEVC extension for circular camera arrangements," in *Proc. 3DTV-Conf., True Vis. Capture, Transmiss. Display 3D Video (3DTV-CON)*, Lisbon, Portugal, Jul. 2015, pp. 1–4.

[45] F. Lewandowski, F. Paluszkiewicz, T. Grajek, and K. Wegner, "Subjective quality assessment methodology for 3D video compression technology," in *Proc. IEEE Int. Conf. Signals Electron. Syst. (ICSES)*, Wrocław, Poland, Sep. 2012, pp. 1–5.

[46] F. Lewandowski, F. Paluszkiewicz, T. Grajek, and K. Wegner, "Methodology for 3D video subjective quality evaluation," *Int. J. Electron. Telecommun.*, vol. 59, no. 1, pp. 25–32, 2013.

[47] K. Wegner, O. Stankiewicz, and J. Siast, *3D-CE2a Results on Nonlinear Depth Representation*, document M25019, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, May 2012.

[48] I. Lim and J. Lee, *Improved Nonlinear Depth Representation*, Standard ISO/IEC JTC 1/SC 29/WG, ITU-T SG 16 WP 3 11, Joint Collaborative Team 3D Video Coding Extension Develop., 2nd Meeting, Shanghai, China, Oct. 2012.

[49] S. Shimizu, S. Sugimoto, H. Kimata, and A. Kojima, "Backward view synthesis prediction using virtual depth map for multiview video plus depth map coding," in *Proc. Vis. Commun. Image Process. (VCIP)*, Kuching, Malaysia, Nov. 2013, pp. 1–6.

[50] *Description of Exploration Experiments in 3D Video*, Standard ISO/IEC JTC1/SC29/WG11 MPEG N9596, Antalya, Turkey, Jan. 2008.

[51] *Description of Core Experiments in 3D Video Coding*, Standard ISO/IEC JTC1/SC29/WG11 MPEG N12561, San Jose, CA, USA, Feb. 2012.

[52] O. Stankiewicz, K. Wegner, and M. Domański, *Nonlinear Depth Representation-Extended Results*, document M23791, ISO/IEC JTC1/SC29/WG11, San Jose, CA, USA, Feb. 2012.

[53] O. Stankiewicz, K. Wegner, and M. Kurc, *3D-AVC-CE3 Results on Nonlinear Depth Representation & Coding*, document M23788, ISO/IEC JTC1/SC29/WG11, San Jose, CA, USA, Feb. 2012.

[54] K. Wegner, O. Stankiewicz, M. Tanimoto, and M. Domański, *Enhanced View Synthesis Reference Software (VSRS) for Free-Viewpoint Television*, Standard ISO/IEC JTC1/SC29/WG11 MPEG2013/M31520, Geneva, Switzerland, Oct. 2013.

[55] M. Domański *et al.*, *Poznań Multiview Video Test Sequences and Camera Parameters*, document M17050, ISO/IEC JTC1/SC29/WG11, Xian, China, Oct. 2009.

[56] J. Zhang, R. Li, H. Li, D. Rusanovskyy, and M. M. Hannuksela, *Ghost Town Fly 3DV Sequence for Purposes of 3DV Standardization*, document M20027, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Mar. 2011.

[57] D. Rusanovskyy, P. Aflaki, and M. M. Hannuksela, *Undo Dancer 3DV Sequence for Purposes of 3DV Standardization*, document M20028, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, 2011.

[58] M. Tanimoto, T. Fujii, and N. Fukushima, *1D Parallel Test Sequences for MPEG-FTV*, document M15378, ISO/IEC JTC1/SC29/WG11, Archamps, France, Apr. 2008.

[59] Y.-S. Ho, E.-K. Lee, and C. Lee, *Video Test Sequence and Camera Parameters*, document M15419, ISO/IEC JTC1/SC29/WG11 MPEG, Archamps, France, Apr. 2008.

[60] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-Curves*, document VCEG-M33, ITU-T SG16, Apr. 2001.

[61] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT.500-13, Jan. 2012.

[62] K. Wegner and O. Stankiewicz, *Results of Subjective Evaluation in 3DV-CE2*, document M25244, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, May 2012.

[63] B. T. Oh, *Cross Check of 3D-CE2.a Results of Poznan University*, document m25056, ISO/IEC JTC1/SC29/WG11, MPEG, Geneva, Switzerland, May 2012.

[64] *Description of Core Experiments in 3D Video Coding*, Standard ISO/IEC JTC1/SC29/WG11 N12561, San Jose, CA, USA, Feb. 2012.

[65] F. Bossen, J. Boyce, K. Suehring, X. Li, and V. Seregin, *JVET Common Test Conditions and Software Reference Configurations for SDR Video*, Standard ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11,JVET-K1010, Joint Video Experts Team, 11th Meeting, Ljubljana, Slovenia, Jul. 2018.

OLGIERD STANKIEWICZ received the M.Sc. degree from the Faculty of Electric Engineering, Poznań University of Technology, in 2006, and the Ph.D. degree from the Faculty of Electronics and Telecommunications, Poznań University of Technology, in 2014, where he is currently an Assistant Professor with the Chair of Multimedia Telecommunications and Microelectronics. He is actively involved in ISO standardization activities, where he contributes to the development of the 3D video coding standards. From 2011 to 2014, he was a Coordinator of the development of MPEG reference software for 3D-video coding standards based on AVC. He actively contributed to JCT-3V, MPEG-I, MPEG Free viewpoint TV, and JPEG-PLENO exploration and standardization activities. He has published over 100 papers (journals, proceedings of international conferences, and also MPEG/JPEG databases) on free view television, depth estimation, view synthesis, and hardware implementation in FPGA. He is a co-inventor of several patents and pending patent applications in European and U.S. patent offices. His professional interests include signal processing, video compression algorithms, computer graphics, and hardware solutions. In 2005, he received the second place from the IEEE Computer Society International Design Competition, Washington, DC, USA.

KRZYSZTOF WEGNER received the M.Sc. degree from the Poznań University of Technology, in 2008, where he is currently pursuing the Ph.D. degree. He is a co-author of several papers on free view television, depth estimation, and view synthesis. His professional interests include video compression in multipoint view systems, depth estimation form stereoscopic images, view synthesis for free view television, and face detection and recognition. He is involved in ISO standardization activities, where he contributes to the development of the future 3D video coding standards. He is currently a Co-Chair of the MPEG FTV Group, which aims at the development of a new generation of coding standards for super-multiview displays and free-viewpoint navigation.

MAREK DOMAŃSKI received the M.S., Ph.D., and Habilitation degrees from the Poznań University of Technology, Poland, in 1978, 1983, and 1990, respectively. Since 1993, he has been a Professor with the Poznań University of Technology, where he currently leads the Chair (Department) of the Multimedia Telecommunications and Microelectronics. Since 2005, he has been the Head of the Polish Delegation to MPEG. He has co-authored highly ranked technology proposals submitted in response to MPEG calls for scalable video compression, in 2004, and 3D video coding, in 2011. He also led the team that developed one of the very first AVC decoders for TV set-top boxes, in 2004, and various AVC, HEVC, and AAC HE codec implementations and improvements. He is an author of three books and over 300 papers in journals and proceedings of international conferences. The contributions were mostly on image, video, and audio compression, virtual navigation, free-viewpoint television, image processing, multimedia systems, 3D video and color image technology, and digital filters and multidimensional signal processing. He is a co-inventor of several patents and pending patent applications in European and U.S. patent offices. He served as a member of various steering, program, and editorial committees of international journals and international conferences. He was the General Chairman/Co-Chairman and the Host of several international conferences: the Picture Coding Symposium (PCS 2012), the IEEE International Conference on Advanced and Signal-based Surveillance (AVSS 2013), the European Signal Processing Conference (EUSIPCO 2007), the 73rd and 112nd Meetings of MPEG, the International Workshop on Signals, Systems and Image Processing (IWSSIP 1997 and 2004), the International Conference Signals and Electronic Systems (ICSES 2004), and others.