IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 15, 2021, accepted December 7, 2021, date of publication January 5, 2022, date of current version January 13, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3140537

Nonlinear Depth Quantization Using Piecewise Linear Scaling for Immersive Video Coding

DOHYEON PARK[1], SUNG-GYUN LIM[1], KWAN-JUNG OH[2], GWANGSOON LEE[2], AND JAE-GON KIM[1], (Member, IEEE)

[1]School of Electronics and Information Engineering, Korea Aerospace University, Goyang 10540, Republic of Korea
[2]Immersive Media Research Section, Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Republic of Korea

Corresponding author: Jae-Gon Kim (jgkim@kau.ac.kr)

This work was supported in part by the Institute of Information and communications Technology Planning and Evaluation (IITP) through the Korea Government (Ministry of Science and ICT, MSIT) under Grant 2017-0-00486, and in part by the IITP through the Korea Government (MSIT) under Grant 2018-0-00207 (Immersive Media Research Laboratory).

ABSTRACT Moving Picture Experts Group (MPEG) is developing a standard for immersive video coding called MPEG Immersive Video (MIV) and is releasing a reference software called Test Model for Immersive Video (TMIV) in the standardization process. The TMIV efficiently compresses an immersive video comprising a set of texture and depth views acquired using multiple cameras within a limited 3D viewing space. Moreover, it affords a rendered view of an arbitrary view position and orientation with six degrees of freedom. However, the existing depth quantization applied to depth atlas in TMIV is insufficient since the reconstructed depth is crucial for achieving the required quality of a rendered viewport. To address this issue, we propose a nonlinear depth quantization method that allocates more codewords to a depth subrange with a higher occurrence of depth values located at edge regions, which are important in terms of the rendered view quality. We implement the proposed nonlinear quantization based on piecewise linear scaling considering the computational complexity and bitstream overhead. The experimental results show that the proposed method yields PSNR-based Bjøntegaard delta rate gains of 5.2% and 4.9% in the end-to-end performance for High- and Low-bitrate ranges, respectively. Moreover, subjective quality improvement is mainly observed at the object boundaries of the rendered viewport. The proposed nonlinear quantization method has been adopted into the TMIV as a candidate standard technology for the next MIV edition.

INDEX TERMS Depth quantization, immersive video coding, MPEG immersive video, piecewise linear scaling, 6DoF video coding, versatile video coding.

I. INTRODUCTION

Recently, with increasing commercial interest in virtual reality applications, 360-degree videos have become popular as a new media type affording an immersive experience [1]–[4]. A 360-degree video, also called omnidirectional video, affords three degrees of freedom (3DoF), allowing viewers to view in all directions from a fixed viewing position. Hence, 3DoF videos do not support motion parallax, where the relative position of objects changes based on the user's location. Therefore, they do not provide complete immersion as they do not respond to a viewer's movements. Six degrees of freedom (6DoF) videos are enhanced 3DoF videos that introduce translational movements with rotations along three axes. They provide viewers with an enhanced immersive visual experience with an interactive parallax feature. In other words, 6DoF videos allow viewers to look around in all directions in a viewing space based on body and head movements. To support motion parallax, the alterations of a view position in a viewing space require the rendering of the virtual view at any intermediate point selected by the viewer. Thus, depth-based rendering is necessary for such immersive experiences. Therefore, a 6DoF video comprises dozens of videos and depths that are simultaneously acquired from multiple positions of a viewing space [5]. Each source view is represented by a sequence of depth and texture frames, as shown in Fig. 1, and with camera parameters to enable three-dimensional reconstruction. These multiple videos of a 6DoF video can be natural sequences captured using a real camera array [6]–[9] or synthetic

The associate editor coordinating the review of this manuscript and approving it for publication was Gulistan Raja.

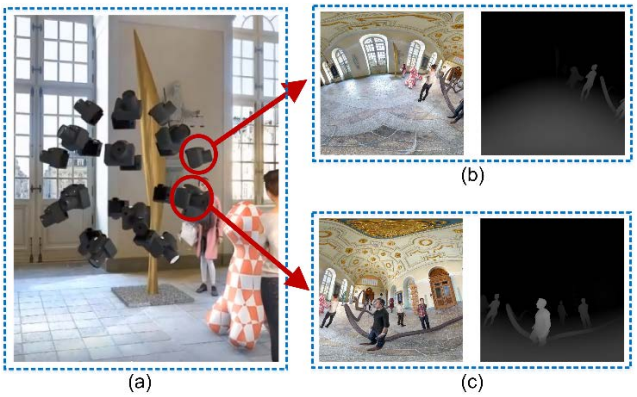

FIGURE 1. Examples of input texture and depth views with correspond-ing viewpoints (Museum sequence). (a) View arrangement, and (b) and (c) first frames of the input views 7 and 11, respectively.

sequences projected by a virtual camera array using computer rendering [10]–[12].

Generally, a 6DoF video that comprises multiview plus depth has a large volume to render high resolutions such as 4K or 8K. Furthermore, the depth information needs to be highly accurate and consistent to achieve a synthesized virtual view with sufficient fidelity. For example, an estimated noisy depth map of natural sequences is significantly inferior to a generated depth map of computer graphics, which makes the synthesis of a virtual view with the required quality difficult [7].

ISO/IEC Moving Picture Experts Group (MPEG) is actively working on standardizing the "Coded Representation of Immersive Media," called MPEG-I [13]. MPEG-I is a project that was initiated at the 116th MPEG meeting in October 2016 and comprises a set of standards including the overall architecture, transmission format, audio/video compression, and metadata for immersive media. The MPEG Immersive Video (MIV), which is the 12th part of MPEG-I, is designed with the capability to compress actual and virtual views captured using multiple cameras with 6DoF within a viewing space [14]. The MIV standardization reached the Final Draft International Standard in July 2021 [15], and further standardization to develop MIV edition 2 is ongoing [16]. At each MPEG meeting, the reference software codec called Test Model for Immersive Video (TMIV) has been released integrating newly adopted tools [17].

The TMIV encoder primarily aims to generate one or more texture and depth atlases by compositing the patches extracted from the input views based on interview redundancy removal. Then, texture and depth atlases are separately encoded using a conventional video codec, such as High Efficiency Video Coding [18] or Versatile Video Coding (VVC) [19]. The TMIV encoder outputs the texture and depth atlases with 10-bit representation and then encodes the atlases using VVenC [20], which is an open-source VVC encoder that inputs 10-bit videos, according to the common test conditions (CTCs) of MIV [21].

In this depth atlas generation process, uniform quantization is applied to convert the 16-bit representation of input source depth values into 10-bit representation values for encoding using VVC. However, the bit-depth scaling of a depth video in TMIV is insufficient as depth is crucial for achieving the required quality of a rendered viewport.

Numerous depth enhancement methods have been actively studied [22]–[24]. However, these studies on pre-processing in advance of video coding do not sufficiently consider the compression efficiency, despite the considerable effect of coding distortion on view synthesis [25]. Therefore, increasing the coding efficiency of depth maps of 6DoF videos is crucial for providing an immersive visual experience. Some studies proposed the reduction of the depth information errors that may occur in the representation of depth with a finite number of bits intended for better coding efficiency [26], [27].

This paper presents a method of depth representation that considers not only a precise expression of the depth value but also the coding efficiency for the depth map. In other words, we present a nonlinear quantization using piecewise linear scaling (PLS) as a normative enhancement method of depth representation in the depth atlas generation inside the MIV standard, which considers the quality of the synthesis and efficiency of the 6DoF video coding. The proposed method aims to improve the end-to-end coding efficiency for depth maps with the limited dynamic range by better utilizing the range and is based on the methodology of Luma Mapping with Chroma Scaling (LMCS) in VVC. LMCS allocates more codewords to a luma range where subjective distortion can easily occur in an input video [29]. The proposed method affords a good trade-off between bit saving and synthesized view quality depending on the depth map reliability in the 6DoF video compression. That is, the depth is accurately represented to increase the rendering quality for computer-generated (CG) contents with high depth quality, and for natural contents (NCs) with low depth quality, the depth is sparsely represented for bitrate reduction. Based on the significant compression efficiency, the proposed method contributed to MPEG [31] has been adopted and integrated into the TMIV [17].

The rest of the paper is organized as follows. Section II presents the overall description of the MIV encoding and the details of the existing depth quantization process in the TMIV encoder. The proposed nonlinear depth quantization using the PLS model is presented in Section III. After the experimental results and the performance analysis of the proposed methods in Section IV, this paper is concluded in Section V.

II. MIV ENCODING AND DEPTH QUANTIZATION

A. MIV ENCODING

Fig. 2 shows the overall architecture of the TMIV encoder. First, source views with different positions and directions are automatically grouped into multiple groups based on a user-specified group number. Then, a view labeling module categorizes each input view within a group into either a basic view or an additional view. The basic views are

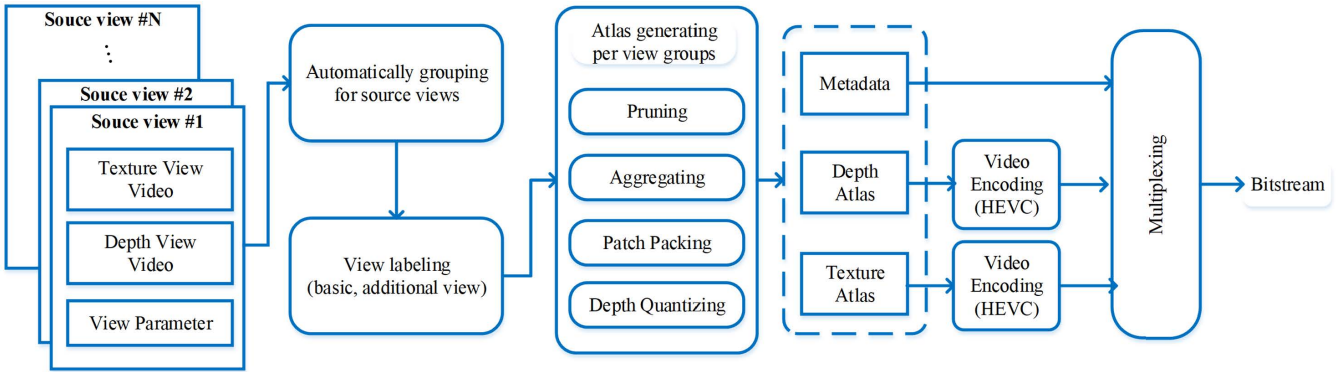

FIGURE 2. Overall architecture of the TMIV group-based encoder.

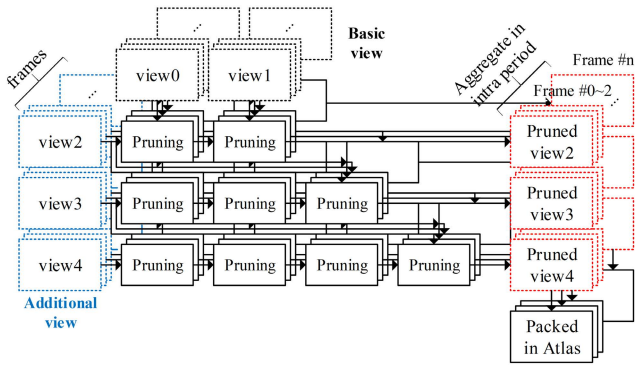

FIGURE 3. Overall process of pruning and aggregating in the atlas generation.

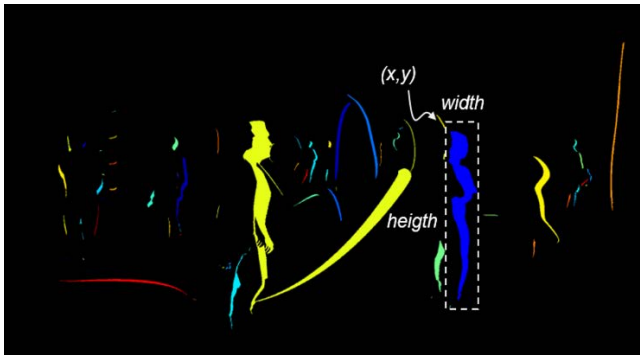

FIGURE 4. Example of an aggregated pruned view and patch information [18].

completely packed into an atlas as a single patch. In contrast, the additional views are pruned by removing inter-view redundancies, and the remaining pixels are clustered and packed into an atlas as multiple patches. In other words, after view categorization, a pruner determines whether individual pixels in each view are removed or preserved based on the inter-view redundancies.

In the pruner, an additional view is projected to the viewpoint of a basic view and the predictable pixels from the basic view are removed as duplicate pixels. After one additional view is pruned, another additional view can be pruned using the basic views and already pruned views. As shown in Fig. 3, the pruning process is applied at the frame level, and the remaining regions are temporally accumulated over all frames belonging to one intra period in the aggregation step. Then, the aggregated regions are clustered into multiple separated regions to be bound by a rectangular box called a patch. Fig. 4 illustrates an example of the clusters obtained by the aggregator [27].

Then, all patches are sequentially packed into single or multiple frames called atlases in the descending order of the patch size. In this process, each patch is packed into an atlas in a raster scan order to occupy an as small as possible packing area while the valid regions are not invaded. Therefore, information for each patch that comprises its top-left corner position, its size with width and height values, and its rotation is included in the metadata to be delivered. Hence, such metadata indicates the view and position from which each patch stemmed when reconstructing additional views in the MIV decoder. Finally, as shown in Fig. 2, video bitstreams of depth and texture atlases are multiplexed with metadata into an MIV output bitstream.

After the generation of the texture and depth atlases, downscaling and quantization are additionally performed on only the depth atlases. The MIV encoder enables resolution reduction by allowing the downscaling of the depth atlas; this is based on the observation that the combination of the resolution reduction and the compression with a low quantization parameter (QP) is efficient for the MIV end-to-end coding performance [32]. Furthermore, depth quantization is performed to efficiently represent the depth values.

B. DEPTH QUANTIZATION

In the MIV standard, a uniform quantization method is adopted for the 10-bit representation of a depth atlas, which may be regarded as an appropriate trade-off between coding efficiency and depth accuracy in terms of the view synthesis quality while accommodating the 10-bit video coding. Accordingly, in the early stage of MIV standardization, a depth atlas is generated based on the normalized disparity

with z_{min} and z_{max} values for efficiently representing the depth information [5]. Hence, the quantized disparity d_q can be obtained as follows:

$$d_q = 2^{bit} \cdot \left(d - \frac{1}{z_{max}} \right) / \left(\frac{1}{z_{min}} - \frac{1}{z_{max}} \right), \quad (1)$$

where d is a disparity value that is an inverse of the real distance. Additionally, z_{min} and z_{max} parameters, which are the minimum and maximum depth range values, respectively, are needed to convert the depth information to the actual distance. They are provided in the view parameter that is input into the MIV encoder with the texture and depth video data. Moreover, z_{min} and z_{max} are the minimum and maximum depth values, respectively, that the camera can capture and not the nearest and farthest distances of a captured object in each view. Therefore, to generate an accurate depth atlas, the quantized disparity can be computed as

$$d_q = 2^{bit} \cdot \frac{d - d_{far}}{d_{near} - d_{far}}, \quad (2)$$

where d_{near} and d_{far} are the disparity values for the nearest and farthest depths found in the input depth views, respectively. They are derived for all frames within each intra period for each view and are included in the metadata and transmitted to the MIV decoder to yield an exact depth value on the decoder side. The normalized disparity, d_q given by (2), for d_{near} and d_{far} is quantized depending on the *bit* value. The *bit* value is determined according to the depth quality, which is assessed based on the first frame of the input depth video in the TMIV [33]. If the depth quality is low, the *bit* value is set to 9 to focus on compression efficiency rather than depth representation accuracy. Otherwise, the value is set to 10 to achieve better view synthesis quality with accurate depth values. Generally, the depth quality is low when the depth is estimated from NCs, whereas it is high when the depth is synthesized in CG sequences.

Moreover, by reserving the range $[0, T]$ of the normalized disparity, the depth atlas contains pixel-wise occupancy information that indicates to the decoder the valid pixels inside the rectangular patches that form the atlas. In other words, if the normalized disparity is less than the guard band value T, the position is an invalid pixel. Thus, occupancy information is embedded in the depth and signaled without a dedicated bitstream. Additionally, the guard band ensures the contour quality of a synthesized view by avoiding noise in zero-valued samples due to coding distortion.

In summary, the normalized disparity of a depth atlas has a value range of $[T, 1023]$ when the depth quality is evaluated as high quality in the depth quality assessment. Otherwise, it has a value range of $[T, 511]$. The depth quantization method described herein is adopted in the first version of the MIV standard [31]. However, the MIV specification allows other methods of depth quantization for the possible enhancement in a switchable manner, and depth quantization methods have been actively studied in the development of the standard [29]–[31].

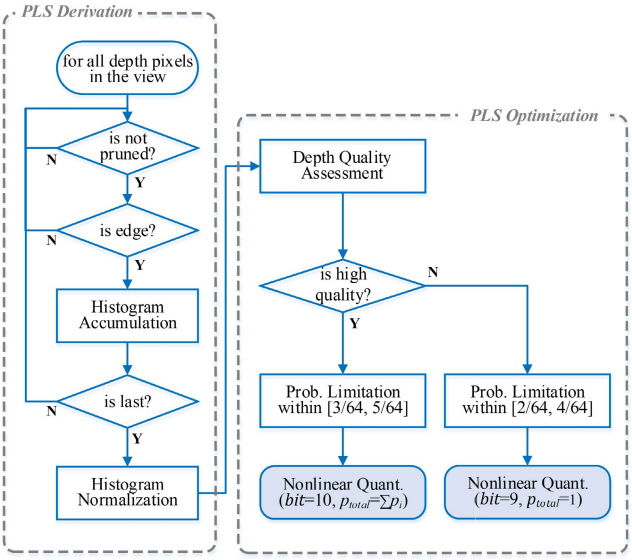

FIGURE 5. Overall flowchart of the proposed nonlinear depth quantization ($N = 16$, $s = 0.25$).

III. PROPOSED NONLINEAR DEPTH QUANTIZATION

The proposed depth quantization is designed to reduce the synthesis view quality degradation due to the coding distortion of the depth atlas while maintaining its bit depth. The proposed nonlinear quantization is performed by adaptively scaling depth subranges according to the importance of each subrange in terms of the rendering quality in a piecewise manner. The degree of the depth importance is computed based on a histogram of the depth values located in the edge regions in the depth view since visual artifacts are likely to appear around an object's edges in the rendered view. The nonlinear quantization employs a PLS because the computational complexity of its nonlinear mapping is low and it does not require a large-sized metadata that should be signaled. Furthermore, the number of nonoverlapping intervals N dividing the entire depth range can be restricted to adjust the encoder/decoder complexity and the metadata signaling overhead. Thus, herein, N is set to 16 as default. Fig. 5 shows the overall flow of the proposed nonlinear depth quantization with a probability limitation parameter s of 0.25, which is detailed in Section III-B. As shown in Fig. 5, the proposed method comprises two parts: the derivation of the PLS model and the optimization of the derived PLS model for integration into the TMIV.

A. DERIVATION OF PIECEWISE LINEAR SCALING MODEL

A PLS model is derived at the encoder side of the TMIV for each view captured from the different camera positions. As mentioned earlier, the proposed method primarily aims to represent foreground object boundaries in the depth atlas with more codewords and background areas with fewer codewords since the quality of the rendered view is likely degraded at foreground object boundaries. Consequently, the proposed method basically allocates more codewords to depth intervals

where important depth pixels occur more frequently using the PLS model.

To derive the PLS model, a depth histogram, which is the occurrence distribution of important depth pixels for each interval bin, is calculated. Depth pixels, which are determined to be in the edge regions via edge detection after the view pruning, are regarded as important; they are used in depth histogram calculations instead of all depth pixels since visual artifacts in the synthesized view are mostly observed in the objects' edges transitioning foreground-to-background [34]–[36].

In more detail, the selection of important depth pixels that are used for calculating a depth histogram comprises two steps. First, depth pixels overlapped by cross-view warping, which projects depth pixels from other views into a current view, are excluded because they are not used for view reconstruction in the MIV decoder. That is, the pruned depth pixels are excluded from the histogram calculation. Then, among the remaining depth pixels after pruning, pixels that are determined to not be in the edges are excluded. Here, edge determination is performed based on the simple gradient calculation. Gradients of the horizontal, vertical, and two diagonal directions are calculated using a one-directional Laplacian operator. The gradient values of g_v, g_h, g_{d1}, and g_{d2} are calculated as follows:

$$g_v = |2 \cdot d\,(i,j) - d\,(i,j-1) - d(i,j+1)|\,, \tag{3}$$

$$g_h = |2 \cdot d\,(i,j) - d\,(i-1,j) - d(i+1,j)|\,, \tag{4}$$

$$g_{d1} = \left| \begin{matrix} 2 \cdot d\,(i,j) - d\,(i-1,j-1) \\ -d(i+1,j+1) \end{matrix} \right|, \tag{5}$$

$$g_{d2} = \left| \begin{matrix} 2 \cdot d\,(i,j) - d\,(i-1,j+1) \\ -d(i+1,j-1) \end{matrix} \right|, \tag{6}$$

where $d\,(i,j)$ indicates the depth value at coordinate (i,j). The maximum and minimum values of the gradients in the horizontal and vertical directions are set as

$$\begin{aligned} g_{v,h}^{min} &= \min\,(g_v, g_h)\,, \\ g_{v,h}^{max} &= \max\,(g_v, g_h). \end{aligned} \tag{7}$$

Similarly, the maximum and minimum values of the diagonal gradients are set as

$$\begin{aligned} g_{d1,d2}^{min} &= \min\,(g_{d1}, g_{d2})\,, \\ g_{d1,d2}^{max} &= \max\,(g_{d1}, g_{d2}). \end{aligned} \tag{8}$$

To determine if a depth pixel corresponds to an edge, the gradient values are compared with each other with the edge strength parameter e_s as

$$\begin{aligned} g_{v,h}^{max} &< e_s \cdot g_{v,h}^{min}\,, \\ g_{d1,d2}^{max} &< e_s \cdot g_{d1,d2}^{min}. \end{aligned} \tag{9}$$

The parameter e_s represents the degree of object boundary strength for determining the depth corresponding to the edge, i.e., more depth samples are detected as object edges with smaller e_s values. The edge strength value e_s is set to 40 according to the experimental results for varying e_s values. The experimental results are described in detail in Section IV-B. If both conditions in (9) are not satisfied, then the depth pixel is excluded from the histogram calculation. The pixels remaining after the above elimination processes are considered as important depth pixels. Subsequently, a normalized histogram is calculated by dividing the range $[d_{far}, d_{near}]$ into N equal intervals and accumulating the important pixels for each interval.

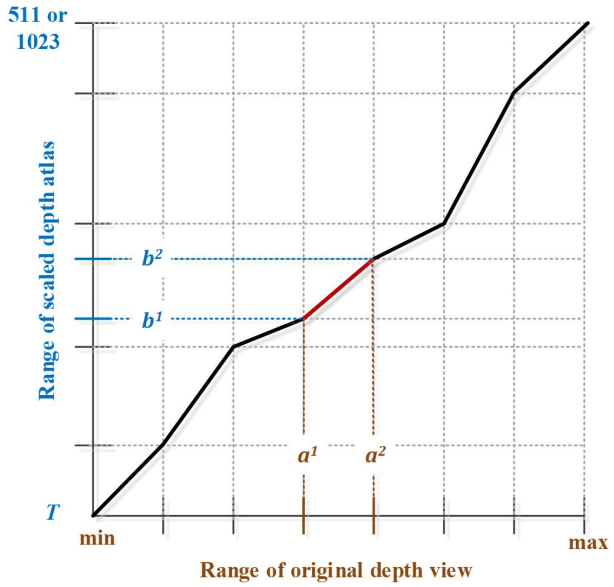

FIGURE 6. Example of the piecewise linear scaling (the number of intervals is 7).

An interval with a high histogram value can be regarded as a more critical range in terms of rendering quality. Hence, each range is scaled up or down by multiplying the corresponding probability of each interval given by the histogram. The proposed depth quantization with the PLS model is

$$d_q = \left(\frac{b_i^2 - b_i^1}{a_i^2 - a_i^1} \right) \cdot \left(d - a_i^1 \right) + b_i^1. \tag{10}$$

Here, d is an original disparity value that is to be scaled, and i is the interval index for the d value. The i-th interval in the original disparity range $(a_i^1, a_i^2]$ is mapped to the i-th interval on the quantized disparity range $(b_i^1, b_i^2]$. As described in Section II-B, the scaled depth atlas is generated as a 9- or 10-bit depth representation based on the depth quality assessment results. As shown in Fig. 6, the original disparity value in each interval is mapped to the corresponding scaled disparity range in a piecewise linear manner.

B. PLS OPTIMIZATION FOR TMIV INTEGRATION

If the proposed depth quantization is applied to the TMIV by directly reflecting the histogram calculated in Section III-A, the contrast of the depth atlas can be significantly highlighted since it allows the areas with lower local contrast to gain a higher contrast. Consequently, the coding performance of

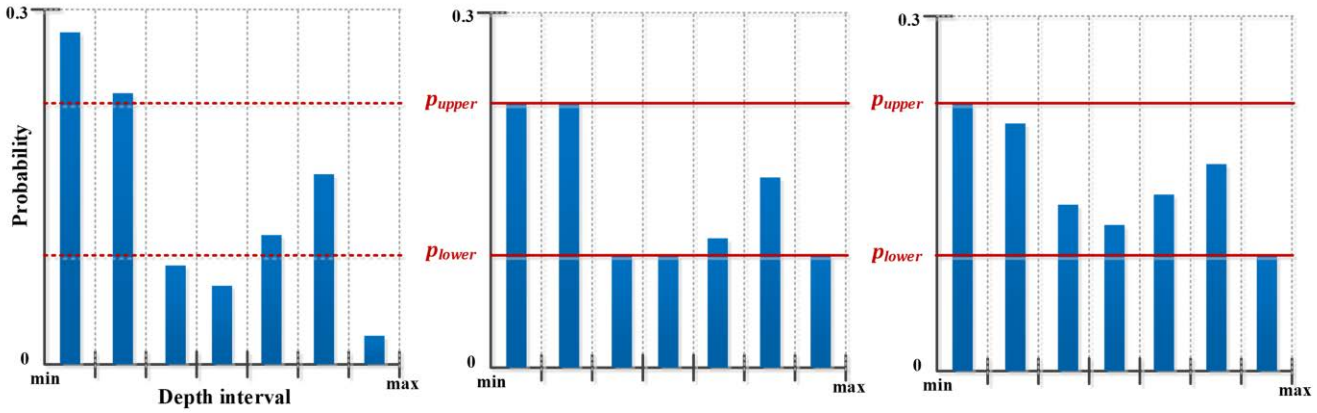

FIGURE 7. Example of the probability range limitation (the number of intervals is 7). The normalized depth histogram (Left) and the results of the probability range limitation with the clipping method (Middle) and scaling method (Right).

the over-scaled depth atlas would be low and the quality of rendered view would be degraded. Therefore, as shown in the model optimization stage in Fig. 5, the derived PLS model is refined herein by adjusting the calculated histogram so that the nonlinear quantization using the PLS model enhances the coding performance while avoiding over-scaling.

First, the interval ranges of the scaled depth are limited to within the predefined allowed range to avoid over-scaling. That is, the depth occurrence probability of each histogram interval is limited to be within the range $[p_{lower}, p_{upper}]$ before the generation of the PLS model. To limit the probability values of all bins to the predefined range, the probability clipping or scaling methods can be used. In the probability clipping method, the probability values that are out of range are mapped to p_{lower} or p_{upper} and those within the range are retained. However, the adjusted probabilities with the clipping method may not maintain the distribution behavior of the histogram. In contrast, in the probability scaling method, the probabilities of all the bins are linearly scaled to the range $[p_{lower}, p_{upper}]$. Therefore, the probability distribution is maintained while limiting the histogram values, unlike in the clipping method. Fig. 7 illustrates examples of the probability adjustment using the probability clipping and scaling methods when the number of depth intervals is 7.

To enable more suitable effects to be mainly obtained by depth quantization based on a given depth view quality, the clipping or scaling probability limitation method is applied with different limits by adjusting the p_{lower} and p_{upper} values. For high depth quality, the probability limitation is applied to retain the range of mapped depth within $\pm(s \times 100)\%$ of the original range, where s is the limitation parameter of the probability value and can be empirically set. Table 1 presents the experimental results for varying s values of {0.15, 0.2, 0.25, 0.3, 0.35} for the optional sequences of ChessPieces, Hijack, and Cadillac in MIV CTCs [21] and for the case without the probability limitation. If the probability limitation is not applied, a significant loss of 22.43% is observed.

TABLE 1. Experimental results of the end-to-end coding performance of the proposed probability limitation method (scaling) with varying s (over TMIV10.0, 17 frames).

Sequence	High-BR BD rate Y-PSNR					
	$s = 0.15$	$s = 0.2$	$s = 0.25$	$s = 0.3$	$s = 0.35$	*No limit*
ChessPieces	−1.80%	−2.30%	−2.50%	−1.30%	−1.70%	21.60%
Hijack	1.00%	1.80%	2.10%	2.70%	3.60%	27.40%
Cadillac	−0.50%	−0.70%	−2.00%	−1.00%	−1.70%	18.30%
Average	**−0.43%**	**−0.40%**	**−0.80%**	**0.13%**	**0.07%**	**22.43%**

Furthermore, the largest average gain of 0.8% is obtained when a probability limitation is applied with an s value of 0.25. If s is 0.25, the scaled depth interval is limited to ±25% of that of the uniform distribution case, which exhibits a histogram value of 1/16 when N is set to 16, by setting the values of p_{lower} and p_{upper} to 3/64 and 5/64, respectively. That is, the scaling slope is determined within 0.75–1.25 according to the corresponding histogram value. In the case of low depth quality, to generate a highly compressible depth atlas, the dynamic range of mapped depth is reduced by setting the range as [2/64, 4/64], which corresponds to the slope of 0.5–1.0.

Finally, in case of high depth quality, the scaled probabilities are renormalized (i.e., p_{total} is set to $\sum p_i$ in (11)) to improve the depth accuracy by retaining the entire dynamic range of the input depth view. Otherwise, the scaled probabilities are retained (p_{total} is set to 1) to reduce the entire dynamic range of the input depth view. Additionally, the representation accuracy of the quantized depth is adjusted according to the depth quality. For sequences with accurate depths, such as a CG video, the rendering quality can be improved through more accurate 10-bit depth representation, while more bit saving in the compression is possible for low-quality depths, such as NCs, with 9-bit representation. The final scaled disparity range $(b_i^1, b_i^2]$ with interval index i can

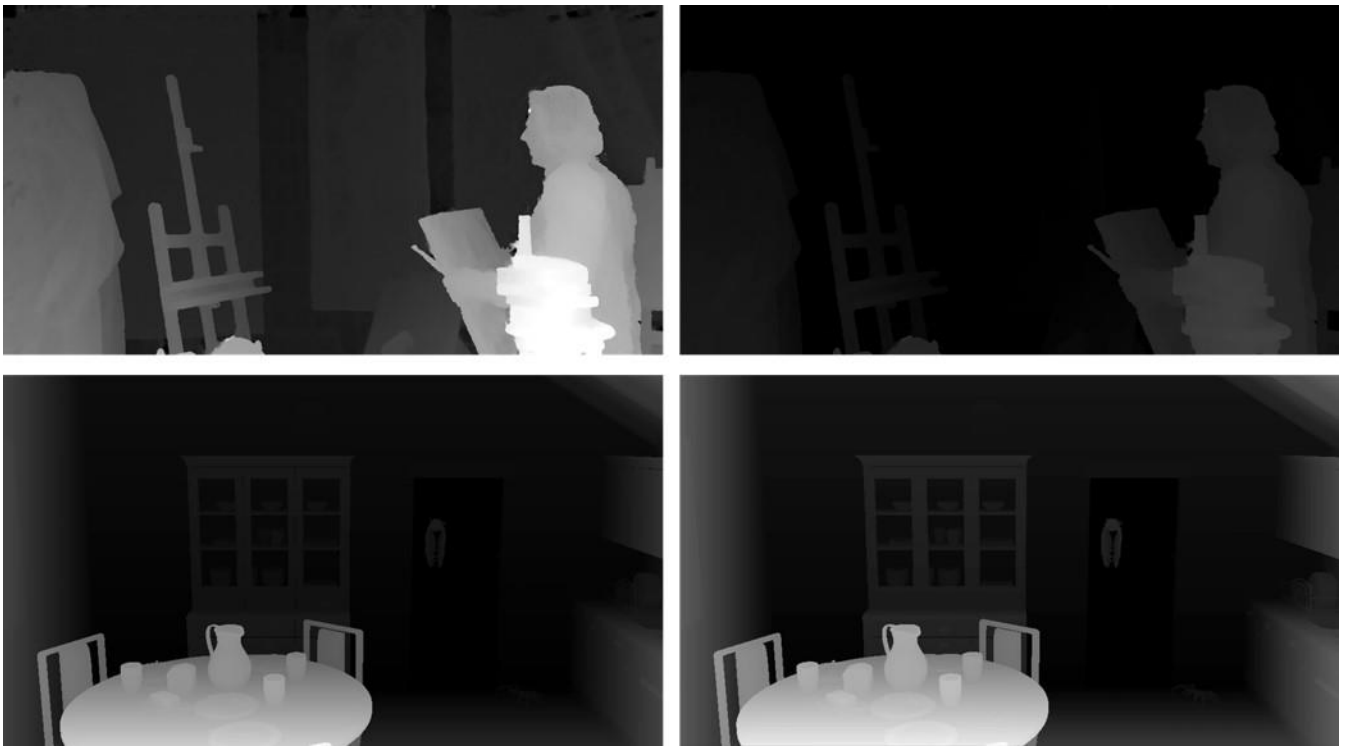

FIGURE 8. Comparison of basic depth views (Left: input depth view, Right: quantized depth view) (Top: Painter, Bottom: Kitchen).

be calculated as

$$b_i^1 = b_{i-1}^2,$$
$$b_i^2 = 2^{bit} \cdot \left(d_{near} - d_{far}\right) \cdot \frac{p_i}{p_{total}} + b_i^1, \tag{11}$$

where b_0^1 is set to 0 and p_i is the scaled probability of the i-th interval.

Furthermore, the generated depth atlas includes an occupancy guard band with the T value described in Section II-B. Fig. 8 compares the input depth view and the corresponding part in the generated depth atlas to which the proposed quantization is applied for NC and CG sequences. For NC sequences (top of Fig. 8), the generated depth atlas looks darker than the original depth view as the entire dynamic range is reduced, which causes bitrate reduction in the compression. For CG sequences, the edges of the depth atlas can be made clearer by heightening the contrast, as shown in the bottom of Fig. 8. Accordingly, depth is more accurately represented as the rendering quality is more effective than bit saving for affording better end-to-end coding performances.

To perform inverse depth quantization in the decoder-side, the number of intervals and the mapped interval boundaries should be signaled in the metadata. To enable the signaling of such metadata for the proposed quantization in a normative way, we propose that syntax structures [31] should be embedded in the bitstream specified in the visual volumetric video-based coding (V3C) standard as an extension for MIV [37]. The V3C standard specifies the bitstream format for transmitting coded volumetric videos, such as an immersive video and point clouds based on video (V-PCC). MIV and V-PCC bitstream syntaxes are aligned in a singled V3C bitstream format because they exhibit high similarity.

Consequently, the probability scaling method affords better coding performance than the probability clipping method; the coding performances are described in detail in Section IV-B. The proposed nonlinear depth quantization with probability scaling has been validated by experts in the MPEG-I Visual group over several meetings with core experiments (CEs) [29]. Finally, the proposed method has been adopted and integrated into the TMIV [17].

IV. EXPERIMENTAL RESULTS

In this section, several experiments are performed to evaluate the performance of the proposed depth quantization in comparison with the current depth quantization of TMIV, which is described in Section II-B. The proposed method was implemented on top of TMIV and evaluated under the MIV CTCs [21]. The test conditions and experimental results of the proposed method are described herein.

A. TEST CONDITIONS

MIV CTCs are established to ensure that all the proposed technologies in the development of the MIV standard employ the same evaluation conditions for fair comparisons. They define the test sequences, specify how anchors are generated,

TABLE 2. Sequence configurations for Texture (T) and Depth (D) video in CTCs.

Test sequence	Content	# views	# Atlases	Atlas Resolution
ClassroomVideo	CG	15	2	T: 4096 × 2176 D: 2045 × 1088
Museum	CG	24	2	T: 2048 × 4352 D: 1024 × 2176
Fan	CG	15	2	T: 1920 × 4640 D: 960 × 2320
Kitchen	CG	25	2	T: 1920 × 4640 D: 960 × 2320
Painter	NC	16	2	T: 2048 × 4352 D: 1024 × 2176
Frog	NC	13	2	T: 1920 × 4640 D: 960 × 2320
Carpark	NC	9	2	T: 1920 × 4640 D: 960 × 2320
Chess	CG	10	2	T: 2048 × 4352 D: 1024 × 2176
Group	CG	21	2	T: 1920 × 4640 D: 960 × 2320

TABLE 3. Experimental results showing the performance enhancement by the refinement with probability scaling over the probability clipping in terms of the BD rate ($s = 0.25$, over TMIV8.0.1, 97 frames).

Sequence	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR
ClassroomVideo	−0.50%	−0.60%	−0.40%	−0.60%
Museum	0.20%	0.30%	−0.10%	−0.10%
Fan	−2.10%	−5.30%	−4.90%	−7.80%
Kitchen	3.10%	1.00%	0.50%	−0.10%
Painter	−11.60%	−14.00%	−11.70%	−14.40%
Frog	−3.60%	−4.90%	−4.20%	−5.50%
Carpark	−7.40%	−9.40%	−8.00%	−10.00%
Chess	0.40%	−0.10%	0.10%	0.00%
Group	2.20%	1.00%	0.90%	0.30%
Average	**−2.10%**	**−3.60%**	**−3.10%**	**−4.20%**

and provide procedures and templates for reporting the experimental results of the contributed technologies. The technical approach to conduct coding experiments is following these steps [21].

1) Compress test sequences based on the TMIV in combination with the VVenC,
2) Synthesize intermediate views from decoded views and metadata,
3) Render viewports of pose traces with a limited movement,
4) Evaluate coding efficiency considering both decoded views and synthesized views.

As shown in Table 2, the test sequence set contains both CG sequences with near-perfect depth views and NC sequences with estimated depth views. The test sequences are captured by various camera arrangements with varying resolutions from Full-HD to 4K. The atlases generated via the MIV encoder are compressed using the VVC open-source software VVenC v0.3.1.0 [20] and VVdeC v1.0.1 [38] with the random-access slow configuration.

Five rate points are specified, corresponding to a group of QPs of {22, 27, 32, 37, 42} for the texture atlas and a group of QPs of {4, 7, 11, 15, 20} for the depth atlas. The QPs for depth are empirically set to low values to maintain the trade-off between rendering quality and coding efficiency. Additionally, two bitrate (BR) modes of High-BR and Low-BR are evaluated to consider diverse bandwidth conditions corresponding to the four lowest QPs and four highest QPs, respectively.

Objective quality is evaluated based on the video quality of all the synthesized views. Two types of metrics are provided to measure the objective quality: peak signal-to-noise ratio (PSNR) for the luminance component (Y-PSNR) and the immersive video PSNR (IV-PSNR) [39]. Y-PSNR is generally used in video coding applications, while IV-PSNR is specifically designed to handle common rendering artifacts that are unnoticeable to human perception. The coding performance is calculated in terms of the Bjøntegaard delta (BD) rate [40] for each metric. The averaged BD rates of each group of four high QPs (High-BR BD rate) and four low QPs (Low-BR BD rate) are reported for each sequence and whole sequences. Moreover, the subjective quality is assessed using a viewport video synthesized along a virtual trajectory predefined by a user as a pose trace [41], [42].

B. EXPERIMENTAL RESULTS

First, the performance of the proposed nonlinear depth quantization with two histogram limitation methods of probability clipping and probability scaling is evaluated in TMIV8.0.1. Table 3 illustrates that the coding performance of the proposed nonlinear quantization with probability scaling is better than that with probability clipping with s of 0.25. Overall, bit savings of 2.1% and 3.6% are observed on average in High- and Low-BR BD rates of Y-PSNR, respectively. Notably, the probability scaling method affords better coding performance by preserving the distribution of the depth histogram. Thus, probability scaling is finally used in the proposed nonlinear depth quantization.

Second, to determine an appropriate value for the object edge detection purpose, the proposed linear scaling method with varying edge strength parameter e_s of {20, 40, 60} in (9) is evaluated, as shown in Fig. 9. For these experiments, the proposed method with the probability scaling was implemented with TMIV10.0 and evaluated according to the MIV CTCs. Table 4 shows the end-to-end coding

FIGURE 9. Object edge detection results according to edge strength parameter e_s (Group sequence, white pixel indicate object edge).

TABLE 4. Experimental results of the end-to-end coding performance of TMIV using the proposed method with different threshold e_s in terms of High-BR BD rate of Y-PSNR ($s = 0.25$, over TMIV10.0, 17 frames).

Sequence	High-BR BD rate Y-PSNR		
	e_s= 20	e_s= 40	e_s= 60
ClassroomVideo	3.60%	3.50%	4.90%
Museum	1.80%	1.60%	1.40%
Fan	−4.60%	−6.10%	−5.40%
Kitchen	−9.80%	−7.00%	−7.80%
Painter	−16.70%	−16.60%	−16.60%
Frog	−5.70%	−6.80%	−6.00%
Carpark	−10.10%	−11.10%	−10.30%
Chess	1.7%	2.1%	1.1%
Group	29.90%	−5.30%	−4.70%
Average	**−1.10%**	**−4.8%**	**−4.8%**

TABLE 5. Experimental results of the end-to-end coding performance of immersive videos with the proposed method in terms of the BD rate ($s = 0.25$, $e_s = 40$, over TMIV10.0, 97 frames).

Sequence	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR
ClassroomVideo	2.2%	2.9%	2.0%	2.7%
Museum	1.6%	1.5%	1.1%	1.5%
Fan	−1.8%	−0.4%	−0.1%	1.3%
Kitchen	−10.9%	−3.4%	−0.6%	1.2%
Painter	−17.5%	−21.1%	−17.6%	−21.9%
Frog	-6.6%	−8.2%	−8.2%	−9.3%
Carpark	−10.1%	−12.7%	−10.8%	−13.5%
Chess	2.1%	0.8%	0.9%	1.3%
Group	−5.9%	−3.7%	−4.1%	−1.9%
Average	**−5.2%**	**−4.9%**	**−4.2%**	**−4.3%**

performance for varying e_s values of 20, 40, and 60 are 1.1%, 4.8%, and 4.8, respectively, in terms of on average High-BR BD rate of Y-PSNR. When e_s is set to 40 or 60, an additional coding gain of 3.7% is observed on average compared with that for the e_s value of 20. Based on the experimental results, robust and accurate edge determination affords improved coding performance, and e_s values of 40 and 60 are appropriate for good end-to-end coding performance.

Table 5 shows the final coding performance of immersive videos with the proposed nonlinear quantization using PLS for depth atlases in comparison to an anchor TMIV, which includes the latest depth quantization method, as described in Section II-B, as a state-of-the-art technology [31]. The proposed method in the experiments is implemented into the TMIV10.0 using the scaling method with an s value of 0.25 when constraining the probability range of the histogram. Additionally, the e_s value for determining the edge depth samples to be accumulated in the histogram is set to 40. Overall, a significant BD rate bit saving of 5.2% and 4.2% are observed on average for the High-BR BD rate of Y-PSNR and IV-PSNR, respectively, compared to the existing quantization method. For NC sequences, the coding gain mainly stems from the dynamic range reduction while the subjective quality is retained. For CG sequences, the subjective quality improves without the bitrate increasing. Furthermore, remarkable performance improvements are noticed in NC sequences, such as Painter [43], Frog [44], and Carpark [45]. Particularly, Painter exhibits the biggest gain of 17.5%. Notably, the proposed method generally performs better at High-BR ranges than at Low-BR ranges.

TABLE 6. Encoding and decoding time for TMIV with the proposed method ($s = 0.25$, $e_s = 40$, over TMIV10.0, 97 frames).

Test Sequence	TMIV Encoding Time (%)	TMIV Decoding Time (%)
ClassroomVideo	100.3%	100.2%
Museum	100.1%	100.0%
Fan	100.1%	100.3%
Kitchen	102.4%	100.3%
Painter	100.0%	100.2%
Frog	100.1%	100.1%
Carpark	100.0%	100.0%
Chess	100.3%	100.0%
Group	100.6%	100.2%
Average	**100.4%**	**100.1%**

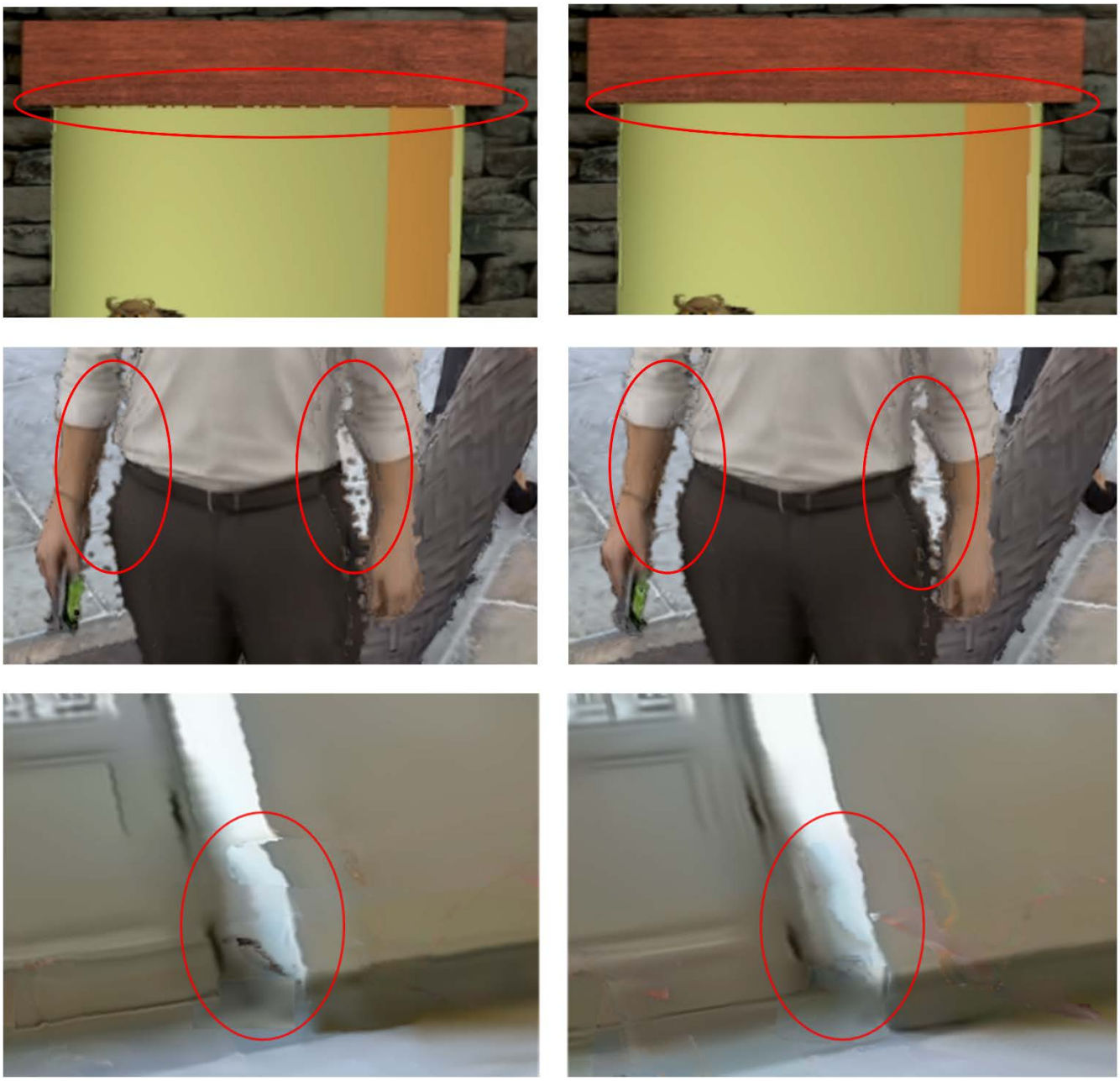

FIGURE 10. Examples of the rendered anchor viewports (Left) and proposed viewports (Right) in the part of Kitchen (Upper) and Group (Middle and Lower) sequences.

Table 6 shows that the proposed quantization slightly changes the complexity of the TMIV with 0.4% and 0.1% increment in the encoding and decoding times, respectively. The TMIV comprises pre-/post-processing for the immersive video and legacy codec for compressing the generated atlases from pre-processing. Although the complexity slightly increases if the proposed method is implemented in the pre-/post-processing of TMIV, it is negligible. Therefore, the proposed depth quantization can be integrated into the TMIV without complexity burden.

Fig. 10 shows the subjective quality comparisons between the viewports rendered by the anchor (TMIV with the existing quantization method) and the proposed method with Kitchen and Group sequences. In Fig. 10, severe artifacts that are shown in anchor viewports are considerably reduced in the viewport rendered with the proposed method (red-colored circles). As shown in the viewport examples, the proposed method significantly improves the visual quality in some sequences and affords considerable BD-rate coding gain.

V. CONCLUSION

This paper presents nonlinear depth quantization method using a PLS model to enhance the immersive video coding performance by efficiently representing the depth information. In the proposed method, the PLS model is derived based on the importance of depth value in terms of rendering quality. The edge depth pixels among the depth pixels remaining after pruning are assumed to be the important ones. Therefore, the mapped range of each depth interval is adaptively determined according to the histogram probability of the important depth values in the corresponding depth interval. In other words, more codewords are allocated to a depth interval with a higher occurrence of edge depth values, which is regarded as an essential range in terms of rendering quality. Furthermore, the PLS model is refined to avoid over-scaling and afford a good trade-off between bit saving and rendering quality according to the depth quality of the given sequence by adjusting the entire mapped depth range.

The experimental results show that the proposed method affords considerable end-to-end coding gain for immersive videos with respect to BD-rate bit saving while maintaining the subjective rendering quality. Consequently, a coding gain of 5.2% average BD rate is observed with noticeable subjective quality enhancement. Furthermore, the proposed method does not change the computational complexities of the TMIV encoder and decoder. After being thoroughly reviewed in several standardization meetings, the proposed method has been integrated into the TMIV and is considered a candidate technology for version 2 of the MIV standard.

REFERENCES

[1] J. Chakareski, "UAV-IoT for next generation virtual reality," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5977–5990, Dec. 2019.

[2] Z. Lai, Y. C. Hu, Y. Cui, L. Sun, N. Dai, and H.-S. Lee, "Furion: Engineering high-quality immersive virtual reality on Today's mobile devices," *IEEE Trans. Mobile Comput.*, vol. 19, no. 7, pp. 1586–1602, Jul. 2020.

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

[4] F. Isgro, E. Trucco, P. Kauff, and O. Schreer, "Three-dimensional image processing in the future of immersive media," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 288–303, Mar. 2004.

[5] 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.

[6] M. Tanimoto, "FTV (free-viewpoint television)," *APSIPA Trans. Signal Inf. Process.*, vol. 1, pp. 1–14, Sep. 2012.

[7] A. Schenkel, D. Bonatto, S. Fachada, H.-L. Guillaume, and G. Lafruit, "Natural scenes datasets for exploration in 6DOF navigation," in *Proc. Int. Conf. 3D Immersion (IC3D)*, Dec. 2018, pp. 1–8.

[8] 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.

[9] P. Goorts, M. Dumont, S. Rogmans, and P. Bekaert, "An end-to-end system for free viewpoint video for smooth camera transitions," in *Proc. Int. Conf. 3D Imag. (IC3D)*, Dec. 2012, pp. 1–7.

[10] T. Senoh, N. Tetsutani, and H. Yasuda, "Depth estimation and view synthesis for immersive media," in *Proc. Int. Conf. 3D Immersion (IC3D)*, Dec. 2018, pp. 1–8.

[11] S. Rogge, D. Bonatto, J. Sancho, R. Salvador, E. Juarez, A. Munteanu, and G. Lafruit, "MPEG-I depth estimation reference software," in *Proc. Int. Conf. 3D Immersion (IC3D)*, Dec. 2019, pp. 1–6.

[12] D. Mieloch, O. Stankiewicz, and M. Domanski, "Depth map estimation for free-viewpoint television and virtual navigation," *IEEE Access*, vol. 8, pp. 5760–5776, 2020.

[13] *Call for Proposals on 3Dof+ Visual*, document ISO/IEC JTC1/SC29/WG11 MPEG/N18709, Jan. 2019.

[14] R. Koenen and M. Champel, *Requirements MPEG-I Phase 1b*, document ISO/IEC JTC1/SC29/WG11 MPEG/N7331, Jan. 2018.

[15] *Text of ISO/IEC FDIS 23090-12 MPEG Immersive Video*, document ISO/IEC JTC1/SC29/WG4/N00111, Jul. 2021.

[16] *Draft Use Case and Requirements for MIV Edition-2*, document ISO/IEC JTC1/SC29/WG2/N0143, Oct. 2021.

[17] *Test Model 10 for MPEG Immersive Video*, document ISO/IEC JTC1/SC29/WG4 N0112, Jul. 2021.

[18] *High Efficiency Video Coding*, Standard ISO/IEC 23008-2, ISO/IEC JTC1/SC29, Apr. 2013.

[19] *Versatile Video Coding*, Standard ISO/IEC 23090-3, ISO/IEC JTC1/SC29, Jul. 2020.

[20] *Fraunhofer HHI VVenC Software Repository*. Accessed: May 2021. [Online]. Available: https://github.com/fraunhoferhhi/vvenc

[21] *Common Test Conditions for MPEG Immersive Video*, document ISO/IEC JTC1/SC29/WG4/N0085, May 2021.

[22] J. Yang, X. Ye, and P. Frossard, "Global auto-regressive depth recovery via iterative non-local filtering," *IEEE Trans. Broadcast.*, vol. 65, no. 1, pp. 123–137, Mar. 2019.

[23] J. Choi, D. Min, and K. Sohn, "Reliability-based multiview depth enhancement considering interview coherence," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 603–616, Apr. 2014.

[24] D. Mieloch, A. Dziembowski, and M. Domanski, "Depth map refinement for immersive video," *IEEE Access*, vol. 9, pp. 10778–10788, 2021.

[25] A. Dziembowski, M. Domanski, A. Grzelka, D. Mieloch, J. Stankowski, and K. Wegner, "The influence of a lossy compression on the quality of estimated depth maps," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, May 2016, pp. 1–4.

[26] O. Stankiewicz, G. Lafruit, and M. Domański, "Multiview video: Acquisition, processing, compression and virtual view rendering," in *Image and Video Processing and Analysis and Computer Vision*, vol. 6, R. Chellappa and S. Theodoridis, Eds. New York, NY, USA: Academic, 2018, pp. 3–74.

[27] J. M. Boyce, R. Dore, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V. K. M. Vadakital, and L. Yu, "MPEG immersive video coding standard," *Proc. IEEE*, vol. 109, no. 9, pp. 1521–1536, Sep. 2021.

[28] T. Lu, F. Pu, P. Yin, S. McCarthy, W. Husak, T. Chen, E. Francois, C. Chevance, F. Hiron, J. Chen, R.-L. Liao, Y. Ye, and J. Luo, "Luma mapping with chroma scaling in versatile video coding," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2020, pp. 193–202.

[29] S.-G Lim, D. Park, J.-G. Kim, J. Jeong, and G. Lee, *[MIV] CE1-Related: Piecewise Linear Scaling of Geometry Atlas*, document ISO/IEC JTC1/SC29/WG4/M56006, Jan. 2021.

[30] S.-G Lim, D. Park, J.-G. Kim, J. Jeong, K.-J. Oh, and G. Lee, *[MIV] CE3.2: Piecewise Linear Scaling of Geometry Atlas*, document ISO/IEC JTC1/SC29/WG4/M56481, Apr. 2021.

[31] S.-G Lim, D. Park, J.-G. Kim, J. Jeong, K.-J. Oh, and G. Lee, *[MIV] CE3.2: Piecewise Linear Scaling of Geometry Atlas*, document ISO/IEC JTC1/SC29/WG4/M57419, Jul. 2021.

[32] B. Sonneveldt and B. Kroon, *Depth-Map Scaling for Pixel-Rate Reduction*, document ISO/IEC JTC1/SC29/WG11 MPEG/M52365, Jan. 2020.

[33] A. Dziembowski, D. Mieloch, M. Domanski, G. Lee, and J. Jeong, *Immersive Video CE1.2: Geometry scaling*, document ISO/IEC JTC1/SC29/WG11 MPEG/M54176, Jun. 2020.

[34] J. Fleureau, R. Doré, F. Thudor, T. Tapie, G. Briand, and B. Chupeau, *Graph-Based Pruning for Natural Contents*, document ISO/IEC JTC1/SC29/WG11 MPEG/M52414, Jan. 2020.

[35] L. Fang, Y. Xiang, N.-M. Cheung, and F. Wu, "Estimation of virtual view synthesis distortion toward virtual view position," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 1961–1976, May 2016.

[36] L. Fang, N.-M. Cheung, D. Tian, A. Vetro, H. Sun, and O. C. Au, "An analytical model for synthesis distortion estimation in 3D video," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 185–199, Jan. 2014.

[37] *Information Technology-Coded Representation of Immersive Media—Part 5: Visual Volumetric Video-Based Coding (V3C) Video-Based Cloud Compression (V-PCC)*, Standard ISO/IEC 23090-5, ISO/IEC JTC1/SC29, Jun. 2021.

[38] *Fraunhofer HHI VVdeC Software Repository*. Accessed: May 2021. [Online]. Available: https://github.com/fraunhoferhhi/vvdec

[39] *Software Manual of IV-PSNR for Immersive Video*, document ISO/IEC JTC1/SC29/WG11 MPEG/N18709, Jul. 2019.
[40] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-Curves*, document VCEG-M33, Mar. 2001.
[41] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document ITU-T Rec. P.910, Apr. 2008.
[42] *Subjective Test Methodologies for 360◦ Video on Head-Mounted Displays*, document ITU-T Rec. P.919, Nov. 2020.
[43] D. Doyen, G. Boisson, and R. Gendrot, *[MPEG-I Visual] New Version of the Pseudo-Rectified Technicolorpainter Content*, document ISO/IEC JTC1/SC29/WG11 MPEG/M43366, Jul. 2018.
[44] B. Salahieh, B. Marvar, M. Nentedem, A. Kumar, V. Popvic, K. Seshadrinathan, O. Nestares, and J. Boyce, *Kermit Test Sequence for Windowed 6DoF Activities*, document ISO/IEC JTC1/SC29/WG11 MPEG/M43748, Jul. 2018.
[45] D. Mieloch, A. Dziembowski, and M. Domanski, *[MPEG-I Visual] Natural Outdoor Test Sequences*, document ISO/IEC JTC1/SC29/WG11 MPEG/M51598, Jan. 2020.

DOHYEON PARK received the B.S. degree in multimedia engineering from Hanbat National University, Daejeon, Republic of Korea, in 2016, and the M.S. degree in electronics and information engineering from Korea Aerospace University, Goyang, Republic of Korea, in 2018, where he is currently pursuing the Ph.D. degree in electronics and information engineering.

He has been involved in several projects focused on immersive video processing and video coding standardization. His current research interests include the development of immersive video processing and video coding techniques based on neural networks.

SUNG-GYUN LIM received the B.S. degree in electronics and information engineering from Korea Aerospace University, Goyang, Republic of Korea, in 2020, where he is currently pursuing the M.S. degree in electronics and information engineering.

He has been involved in several projects focused on immersive video processing and video coding standardization. His current research interests include the development of immersive video processing and video coding techniques.

KWAN-JUNG OH received the B.S. degree in electronic computer engineering from Chonnam University, Gwangju, Republic of Korea, in 2002, and the M.S. and Ph.D. degrees in information and communications engineering from the Gwangju Institute of Science and Technology (GIST), Gwangju, in 2005 and 2010, respectively. From 2010 to 2013, he was with the Samsung Advanced Institute of Technology (SAIT), where he was involved in standardization activities of 3D video coding. He joined the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea, in 2013, where he is currently a Senior Researcher with the Broadcasting and Media Research Laboratory. His research interests include 3D image and video coding, immersive media, and holography.

GWANGSOON LEE received the Ph.D. degree in electronics engineering from Kyungpook National University, Daegu, South Korea, in 2004. He joined the Electronics and Telecommunications Research Institute (ETRI), Daejeon, in 2001. He is currently a Principal Researcher with Realistic-Media Research Sector. He has been involved in immersive video coding standards in ISO/IEC MPEG. His research interests include the immersive video processing, light-field imaging systems, and the three-dimensional video systems.

JAE-GON KIM (Member, IEEE) received the B.S. degree in electronics engineering from Kyungpook National University, Daegu, South Korea, in 1990, and the M.S. and Ph.D. degrees in electronical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1992 and 2005, respectively. From 1992 to 2007, he was with the Electronics and Telecommunications Research Institute (ETRI), where he was involved in the development of digital broadcasting media services, MPEG-2/4/7/21 standards and related applications, and convergence media technologies. From 2001 to 2002, he was a Staff Associate with the Department of Electrical Engineering, Columbia University, New York, USA. Since 2007, he has been with Korea Aerospace University, Goyang, South Korea, where he is currently a Professor with the School of Electronics and Information Engineering. From 2014 to 2015, he was a Visiting Scholar with the Video Signal Processing Laboratory, University of California, San Diego. He has been involved in video coding standards in JCT-VC and JVET activities of ITU-T VCEG and ISO/IEC MPEG. His research interests include image/video compression, video signaling processing, and immersive video.