

V-PCC: Performance Evaluation of the First MPEG Point Cloud Codec

By Céline Guede, Pierre Andrivon, Jean-Eudes Marvie, Julien Ricard, Bill Redmann, and Jean-Claude Chevet

Abstract

Representation of 3D scenes has become increasingly important in several industries using virtual and augmented reality technologies. The point cloud format is well suited for such representations. Indeed, point clouds can be created with a simple capture process and modest processing, enabling a realtime, end-to-end point cloud distribution chain. However, point cloud compression is required to obtain data rates and file sizes that could be economically viable for the industry. Standardization is required to ensure interoperability. In 2021, the International Organization for Standardization (ISO) Moving Picture Experts Group (MPEG) plans to publish a standard for its first point cloud codec, MPEG-I Part 5: Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC). This standard enables a world of new services and applications, including cultural heritage, telepresence, and new forms of entertainment. In this paper, we review the principal use cases targeted by the V-PCC standard. We present the architecture of the V-PCC codec and describe its main tools, by giving insights into complexity at the encoder and decoder levels and explaining profiles and conformance points in V-PCC. We then present the methodology established, as a collaboration between industry and academics, for the evaluation of the V-PCC codec performance and the methodology's origins. This methodology was applied to the MPEG point cloud compression test model software (named TMC2) to consistently evaluate technologies proposed during the standardization process. Finally, we compare the performance of the main V-PCC tools available for lossy and lossless compression. Finally, the conclusion provides elements in favor of a near-term deployment of V-PCC in the media industry.

Keywords

Immersive video coding, performance evaluation, point cloud compression

Introduction

Advanced 3D representations of the world enable more immersive forms of interaction and communication and allow machines to understand, interpret, and navigate our world. 3D point clouds have emerged as an enabling representation for such information.

A point cloud is a set of points in a 3D space, each with associated attributes, for example, color and material properties. Point clouds can be used to reconstruct an object or a scene as a composition of such points, can be captured using multiple cameras and depth sensors in various setups, and can be made up of thousands up to billions of points to realistically represent reconstructed scenes.

A point cloud is a set of points in a 3D space, each with associated attributes, for example, color and material properties. Point clouds can be used to reconstruct an object or a scene as a composition of such points, can be captured using multiple cameras and depth sensors in various setups, and can be made up of thousands up to billions of points to realistically represent reconstructed scenes.

Compression technologies are needed to reduce the amount of data required to represent a point cloud. As such, technologies supplying lossy but efficient compression of point clouds are needed for use in realtime communications. In addition, technology is sought for lossless point cloud compression in the context of dynamic mapping for autonomous driving, cultural heritage applications, and so on.

Several use cases associated with point cloud data have been identified, and corresponding requirements for point cloud representation and compression have been developed inside the Moving Picture Experts Group (MPEG) community. These will be described in the section “Presentation of the V-PCC Codec.”

The International Organization for Standardization (ISO) MPEG standards addresses the compression of geometry and attributes, such as colors and reflectance,



FIGURE 1. Combination of V-PCC tools in lossy configurations. From left to right: original uncompressed point cloud, 1-Map no enhancement (BO_MC1_RDP0_GS0), 1-Map with PLR and PBF (E2_MC1_PLR_RDP1_PBF). Magnified frames show the associated artifacts, especially on faces and hands. Basketball test content frame #1 (see Ref. 1).

scalable/progressive coding, coding of sequences of point clouds captured over time, and random access to subsets of the point cloud. The acquisition of point clouds is outside of the scope of this standard.

In 2021, ISO/MPEG plans to publish two standards for point cloud compression: Video-based Point Cloud Compression (V-PCC) and Geometry-based Point Cloud Compression (G-PCC) documented, respectively, in MPEG-I part 5 V-PCC and MPEG-I part 9 G-PCC of ISO/IEC JTC1/SC29 WG11.

This paper focuses on the V-PCC standard in both lossless and lossy configurations (**Fig. 1**). An initial V-PCC codec architecture was submitted in response to the call for proposals in October 2017. This architecture was enriched by tools that improve coding performance or minimize 3D object reconstruction artifacts, thereby improving visual quality.

We describe the V-PCC codec architecture and provide an overview of its main coding tools in the section

“Presentation of the V-PCC Codec.” We then present their performance and explain the methodology used for their comparison in the section “Experimental Results.” Finally, short-term deployment recommendations for the media industry are given in the section “Presentation of the V-PCC Codec.”

Presentation of the V-PCC Codec

The democratization of 3D sensors, capable of capturing animated high-definition representations of the world, induces strong needs with respect to applications providing realtime visualization of immersive videos.

For V-PCC, applications have been identified by the MPEG community in Ref. 2, including realtime immersive content viewing with interactive parallax for telepresence, augmented reality (AR) and virtual reality (VR), and 3D free viewpoint video, both shown in **Fig. 2**.

3D immersive telepresence implies realtime communication and preserves as much realism as possible. Such an application requires a bitrate in the range of medium to low realtime processing, with low latency (encoding, decoding, and rendering) and error resilience. 3D broadcasting of sports is another application for point clouds in which the user can replay a part of the game from a preferred point of view. In this case, interoperability between manufacturers is important, and low delay for encoding and decoding must be available.

The MPEG PCC standardization community defined three categories of point cloud test data to address all use cases identified in Refs. 3 and 4: *static*, *dynamic*, and *dynamically acquired*. The V-PCC codec, described in this article, handles the *dynamic* category, which corresponds to dense point cloud data sets varying in time.



FIGURE 2. Point cloud use cases: (on the left) VR/AR telepresence, InterDigital source; (on the right) 3D free viewpoint sport replay broadcasting, Intel source.

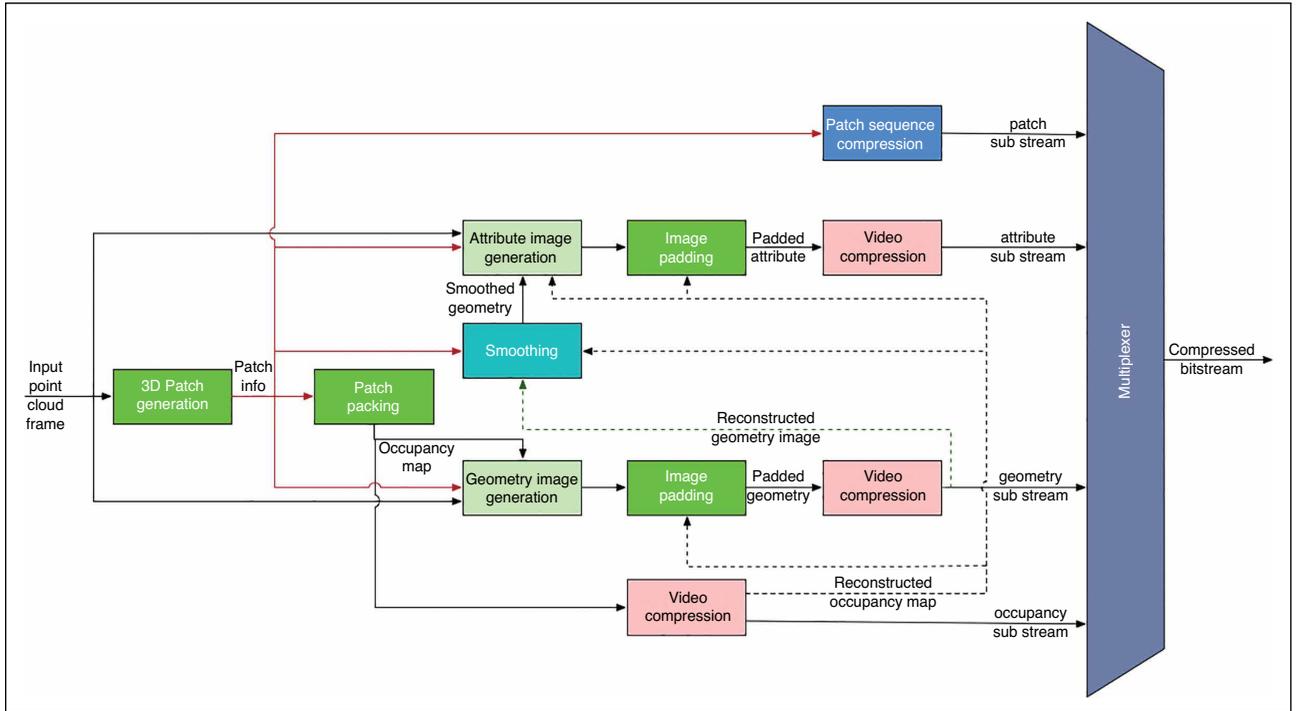


FIGURE 3. V-PCC TMC2 encoding structure from Ref. 6.

Architecture of the V-PCC Codec

The core encoding and decoding processes for V-PCC were inherited from the solution that demonstrated the highest compression efficiency among all submissions, as agreed during the 119th MPEG meeting (Macau). The V-PCC solution is agnostic with respect to which 2D video codec is used. Hence, it can naturally leverage video codec evolution to take advantage of improvements in encoding techniques and performance. The test procedure described in this document leverages the state-of-the-art high efficiency video coding (HEVC) reference model (HM)⁵ implementation for video-based coding.

Figure 3 shows the block diagram of the V-PCC process for encoding dynamic point clouds with color attributes, while the decoding diagram is presented in Fig. 4.

Encoding

In this section, we describe each step of the encoding process, as depicted in Fig. 3. At the encoding input point, each point cloud frame is processed as described hereafter.

“3D patch generation” is performed first (Fig. 3), in which the input point cloud is represented as a set of orthogonal 3D projections onto the six faces of a rectangular parallelepiped (an axis-aligned bounding box of the 3D object). These projections are made of connected components (CC), where each CC is a set of neighboring points having similar normals, as shown in Fig. 5(a).

Each CC is projected onto one of the six bounding box faces, parallel to the main planes XY , XZ , or YZ , by choosing which has a normal closer to the average normal of the points in the CC. The orthographic projection of the geometric information permits us to maintain the distance

of each point relative to the selected face. For instance, a point $p = (x, y, z)$ of a CC that is projected to the XY -plane would result in its value of z being projected to and stored in the (x, y) pixel of that XY -plane.

Optionally, 12 additional projection planes with different orientations (see Ref. 7) can improve the visual quality of the reconstructed point cloud by capturing more points during the projection phase. These additional planes are at 45° along each edge of the bounding box and bring the total number of projections to 18.

Depending on the distribution of points in the CC, more than one value may be projected onto the same coordinate of the projection plane. A trivial approach would simply maintain the value corresponding to the closest point (i.e., the smallest depth value), but this may not allow capturing more complex 3D point distributions (like folds in clothing). This trivial approach, called the *1-Map configuration*, supplies one map for the geometry and one for the attribute. A more sophisticated approach is to project several times and generate several maps for the geometry and corresponding maps for the attributes. To minimize the decoder complexity for applications with low computation capabilities, the V-PCC basic profile limits the number of maps to no more than 2. When a limited number of maps are maintained, some points of the original point cloud may not be projected. This will cause occlusions inside the reconstructed point cloud. However, V-PCC extended profile permits the use of a higher number of maps, for instance, for applications requiring a higher fidelity. Keeping two depth values per coordinate of the projection plane produces the *2-Map mode*. Although the 1-Map case only retains the nearest

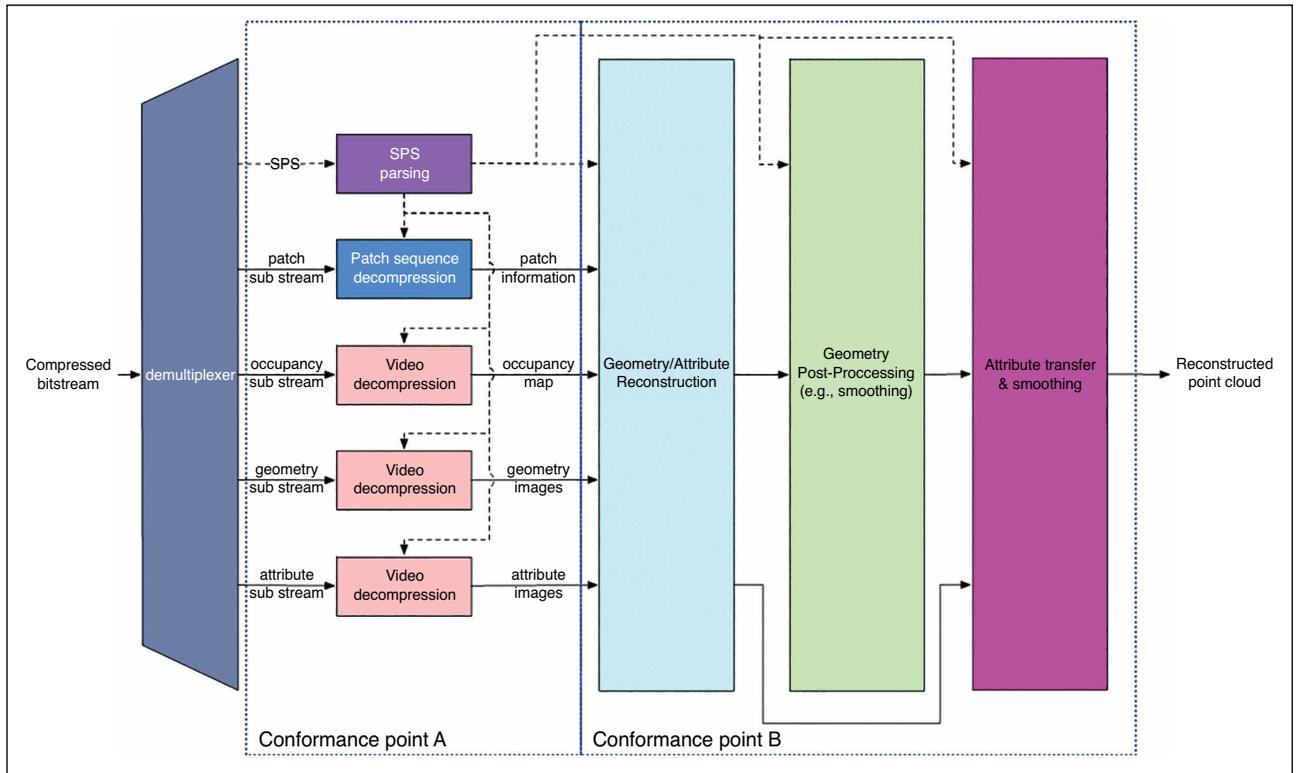


FIGURE 4. V-PCC TMC2 decoding structure from Ref. 6. Conformance points A and B from Ref. 8.

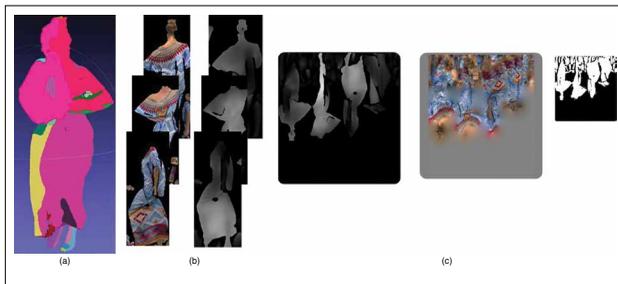


FIGURE 5. V-PCC encoder main steps.

(smallest) depth value, the 2-Map case retains the farthest (largest) depth value. The 2-Map mode better captures the distribution of points in 3D space but at the expense of increasing the amount of projected data to be encoded. The projection step is shown in Fig. 5(b).

The “patch-packing” process (Fig. 3) groups all the previously identified projected areas, also called *patches*, onto a 2D frame and fills this 2D frame in an optimized way (e.g., by minimizing the empty part and ensuring a temporal consistency), making sure that each patch-packing block (e.g., 16×16 block) of the frame contains, at most, a single patch.

The result is called the *atlas* [Fig. 5(c)], a trio of 2D frames constructed according to this projection scheme:

- a geometry frame (“geometry image generation” in Fig. 3) that stores the CC depth values,
- an attribute frame (“attribute image generation” in Fig. 3) that stores the color components (i.e., the texture), and

- an occupancy map, a binary image that indicates which parts of the geometry and texture frames are to be used for the reconstruction stage.

An “image-padding” process (Fig. 3) fills the empty space between patches using the reconstructed occupancy map, with the goal to make the frame better suited for video coding [Fig. 5(c)]. It is used on the geometry and attribute frames.

For lossy configurations, the occupancy map may be downscaled to reduce the number of bits needed to encode the occupancy map. For lossless encoding, the occupancy map must be kept at full resolution for the regeneration of the final point cloud [Fig. 5(c)].

The three 2D frames of each atlas (geometry, attributes, and occupancy) obtained through this process are encoded as three separate videos (SVs), using traditional 2D video compression solutions, such as HEVC or versatile video coding (VVC) (Fig. 3). V-PCC takes advantage of the temporal inter-prediction mechanism of such underlying 2D video encoders for coding temporally stable atlases. Some auxiliary data used for the reconstruction, including patch information, is also entropy-coded.

Alternatively, the reconstruction of the encoded geometry component undergoes a “smoothing” process (Fig. 3) to remove visible artifacts due to lossy coding (mainly due to the downscale of the occupancy map and the coding compression artifacts). To get a better attribute representation, the smoothed geometry and the reconstructed occupancy map are used to build the attribute image.

At the end of the process, the SV and auxiliary patch data bitstreams are multiplexed into the output compressed binary V-PCC bitstream (**Fig. 3**).

Decoding

The decoding process (**Fig. 4**) is the functional inverse of the encoding process. It starts from demultiplexing the input compressed binary bitstream into geometry, attribute, occupancy map, and auxiliary information streams.

The auxiliary information stream, containing information for reconstruction, is decoded. The occupancy map is decoded and, if downsampled on the encoder side, is upsampled to its nominal resolution. The geometry stream is decoded and, in combination with the occupancy map and the auxiliary information, the point cloud geometry information is reconstructed and optionally smoothed along geometric patch boundaries.

Finally, the attributes (e.g., colors) of the point cloud are reconstructed based on the decoded attribute video stream, the reconstructed information for the geometry (which may be smoothed), the occupancy map, and the auxiliary information. Afterward, an additional attribute smoothing method may be used as a final point cloud refinement.

Presentation of V-PCC Main Coding Tools With Complexity Considerations

Many V-PCC coding tools are described in detail in Ref. 6. For this paper, the focus is on tools that visually and objectively impact the decoding and reconstruction of the point cloud in terms of objective performance metrics (refer to conformance point B described in **Fig. 4**). The following is a list of these selected tools with a brief description.

Multiple Maps (Named MC for Map Count)

As mentioned in the section “Architecture of the V-PCC Codec,” there is a 2-Map mode defined in V-PCC. In this configuration, maps keep both the nearest and farthest depth, spaced by at most a surface thickness. This surface thickness is defined on the encoder side and can be adjusted according to the reference source content (i.e., the original and uncompressed point cloud). It aids capture of opposed surfaces, such as the screen and the back of a tablet, without mixing them into the same CC. This multiple-maps concept allows storage of overlapped points and better preserves the point geometries and occlusions in the 3D space.

When the 2-Map mode is set, the input point cloud frame is encoded with two maps for the geometry and two corresponding maps for attributes (i.e., texture). This implies a doubled memory consumption and requires synchronization, since occupancy has one frame for each of the two frames for each of geometry and texture. This could represent a limitation in some low-end devices.

The main syntax element for this tool in the standard⁸ is `vps_map_count_minus1[atlas_idx]`.

Single-Pixel deInterleaving

Single-Pixel deInterleaving (SPI), detailed in Ref. 9, represents geometry frames for two maps in a single geometry frame (and likewise for texture) by interleaving values (in a quincunx shape) of the near and far depth maps. It is represented in **Fig. 6** as the “interleaved depth frame.”

Missing values can be deduced from the neighborhood on the decoder side. The process is performed in 2D and is only used for lossy configurations. For the decoder, this implies parsing the `asps_pixel_deinterleaving_enabled_flag`

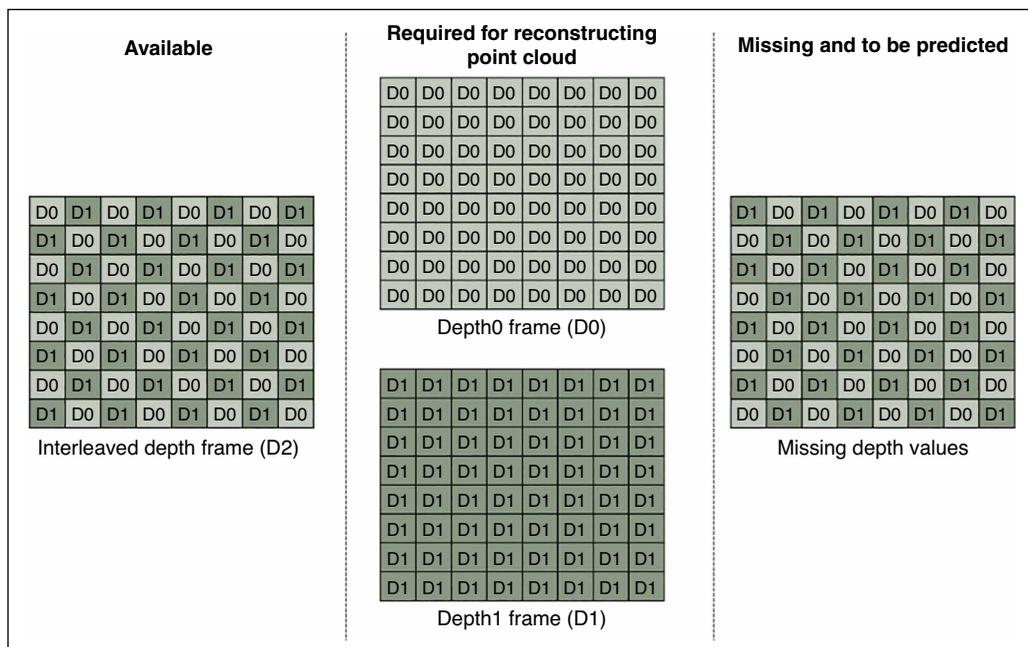


FIGURE 6. Example of geometry de-interleaving process and prediction.

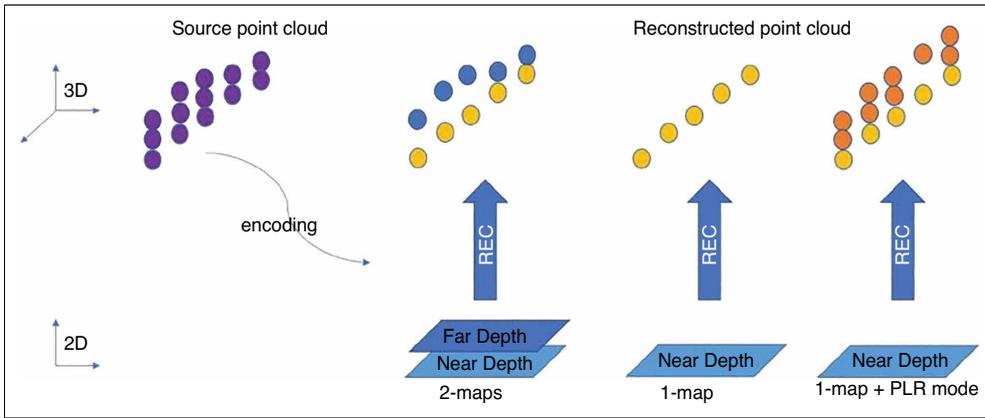


FIGURE 7. Point cloud reconstruction from left to right: 2-Map mode, 1-Map, and PLR approaches.

(see Ref. 8), de-interleaving, and then upscaling the geometry and attribute information for each point cloud frame. Compared to a pure single-map representation, pixel interleaving enhances quality while preserving the performance advantage of decoding only one map.

Point Local Reconstruction

The Point Local Reconstruction (PLR) tool, detailed in Ref. 10, improves the geometry of a map by adding points during the reconstruction. Compared to SPI, which builds interleaved maps, PLR is based directly on the near map, as described in the section “Encoding.” In addition, a prediction mode is defined to be either block by block or patch by patch (with the best mode determined by a rate distortion optimization method in the encoder) and is signaled in the stream to allow the decoder to generate points from the single map received. **Figure 7** shows an input point cloud (in purple) that is represented in 2D with three different approaches: 2-Map mode, 1-Map mode, and 1-Map augmented with PLR. The 2-Map mode reconstructs points in 3D that correspond to the far (blue points) and near (yellow points) map. The 1-Map approach reconstructs only points of the near map (yellow points). PLR conveniently reconstructs the points of

the near map (yellow points), but also points that could have been encoded for the far map (orange points) so as to be as faithful as possible to the reference source point cloud. The PLR process is performed in 2D and is only used in lossy configurations. The attributes for the new points are derived from the attributes of neighboring points. As with SPI, PLR enhances quality while preserving the performance advantage of decoding only one map. It is noted that, like SPI, PLR may also be used on several maps.

The main syntax element for PLR is `asps_plr_enabled_flag`. A small amount of data parsing is needed on the decoder side to know which mode has to be applied by the geometry reconstruction block, among those described in Ref. 10.

Geometry Smoothing

Geometry Smoothing (GS), as detailed in Ref. 11, filters the points at the boundary of patches. For each point identified as belonging to the patch boundaries, highlighted in red in **Fig. 8**, it computes a centroid of the decoded points in a small 3D grid (**Fig. 8**, right). After the centroid and the number of points in the $2 \times 2 \times 2$ grid is derived, a commonly used trilinear filter is applied. This tool is only used for lossy configurations, since its purpose

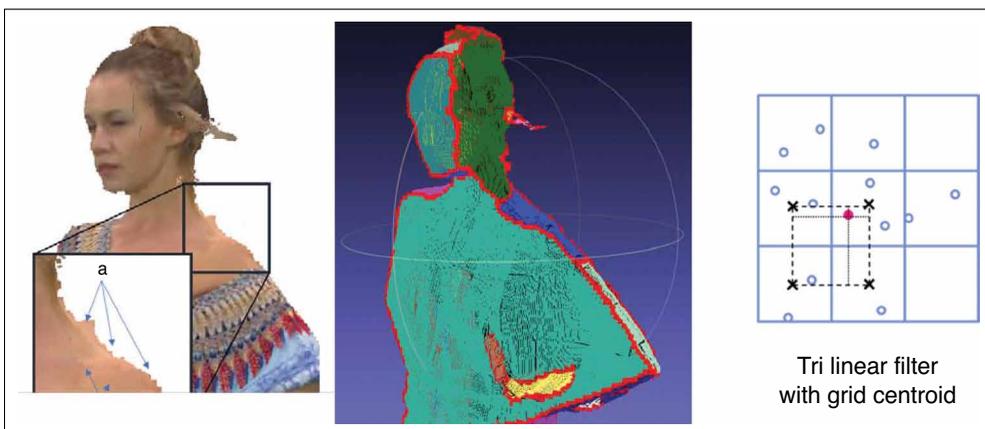


FIGURE 8. Points for GS and trilinear filter from Ref. 6. Left: geometric (a) and photometric (b) artifacts due to occupancy downsampling. Center: patch boundaries are set in red where the GS is applied. Right: the computation of the smoothed point, called centroid (in red), computed from nearest neighbors.

is to correct the artifacts introduced by downsampling the occupancy map described in the section “Encoding.” The section “Comparisons Regarding the Patch Filtering Process” shows the visual quality improvement provided by this tool. However, this is a multi-pass process operating in 3D space, and the multiple passes over the reconstructed points may not be appropriate for devices having limited processing capabilities.

To be applied on the decoder side, the corresponding supplemental enhancement information (SEI) message, `geometry_smoothing` (payloadSize), must be supplied with `gs_method_type[k]` set to 1.

Color Grid Smoothing

The Color Grid Smoothing (CGS), detailed in Ref. 12, aims at averaging potential artifacts in the color values near patch boundaries. The process is done in 3D space so that the correct neighborhood is employed when producing the smoothed attribute value. This tool is only for lossy configurations and may be complementary to the geometry or occupancy map-filtering techniques.

Information on attribute smoothing is transmitted as an SEI message, `attribute_smoothing` (payloadSize).

Patch Border Filtering

As explained in the two previous sections, smoothing patch borders removes some artifacts due to the occupancy map upsampling and video compression, by moving the reconstructed points. The Patch Border Filtering (PBF) process, as detailed in Ref. 13, proposes an alternative way to correct the occupancy map before the reconstruction step. It has the advantage of being performed in 2D. For each patch, the borders of all adjacent patches are projected in the 2D space to adjust the frontier of the patch (**Fig. 9**). This process removes the border artifacts, ensures the alignment of two adjacent patch borders, and avoids holes and any overlapping of points. As with GS, this is used only in lossy configurations by design.

To be applied on the decoder side, the corresponding SEI message, `occupancy_synthesis` (payloadSize) must be set with `os_method_type[k]` set to 1.

Remove Duplicate Points

By default, the reconstruction process creates one 3D point per map for each non-zero value of the occupancy

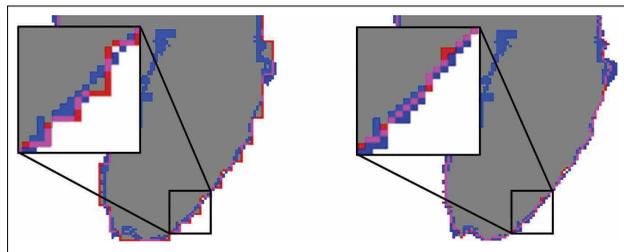


FIGURE 9. Example of patch pictures. The figures show how points are moved: patch points (grey), adjacent patch points (blue), and current patch border [red or purple (when above blue)] before patch filtering (left) and after (right).

map, without looking at the coordinates of the points created. If the number of maps per atlas is greater than 1, then wherever two depth values stored at the same coordinates in two frames are equal, the reconstruction process will create two 3D points having the same coordinates. The process of removing duplicate points, as detailed in Ref. 14, compares the pixel values in depth maps during the reconstruction process and does not create a second point if a pixel value in the far map is equal to the pixel value of the nearer map.

By reducing the number of points that must be treated, this filtering has a positive impact on the complexity and memory consumption of subsequent processes.

Multiple Streams

The Multiple Stream (MS) tool indicates the number of video streams used to encode each of the geometry and the attribute frames. When this tool is activated, the geometry is composed of as many streams as the number of maps (typically two), likewise for the attribute. Each supplementary stream can be, in that case, coded as a delta value compared with the first stream (i.e., the difference between the far and near depth) or as an absolute value. This is detailed in Refs. 10 and 12.

When the MS tool is active, it is better to use the delta coding mode for coding geometry and attribute information, as shown in the section “Lossy 2-Map Results.” However, when the MS tool is deactivated, the behavior is that of absolute coding. Indeed, in that case, the far depth frame has the same characteristics as the nearest depth frame, and the coding becomes more suitable to be compressed as it uses inter-frame prediction supported by 2D video codecs.

This tool is signaled in the bitstream with `vps_multiple_map_streams_present_flag`. Some comparison results are given in the section “Lossy 2-Map Results.”

Raw

The Raw tool, as detailed in Ref. 15, allows coding of any remaining points that have not been gathered into a CC and projected in a specific patch. Instead, these points are aggregated into an identified “raw patch” for which the flag `asps_raw_patch_enabled_flag` signals the presence. An SV flag, `vps_auxiliary_video_present_flag[atlasIdx]`, when set to 1, indicates that raw coded geometry and attribute information for the atlas may be stored in an SV stream (referenced as auxiliary video in the standard). When this flag is set to 0, then raw point patches are present in the same geometry and attribute video streams as the other patch types, as shown in **Fig. 10**. This tool can be used in lossy configurations, but it is mandatory for lossless operations, so as to reflect all the information from the reference source point cloud concerning the otherwise unallied points. The quality gain provided by this Raw information comes at the expense of additional coding.

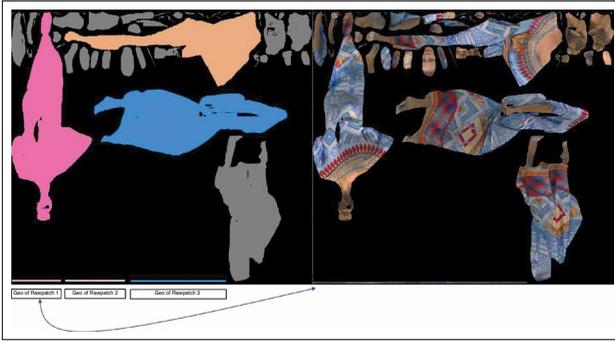


FIGURE 10. Raw points are typically coded below other projected patches. Example for a geometry map using false colors for illustrative purpose (on the left) and for an attribute map with frame padding disabled (on the right).

Enhanced Occupancy Map

This tool, as detailed in Ref. 16, proposes to handle “in-between points” that lie from the near and possibly up to the far maps (when it exists) by using a coded bit sequence (or codeword) in which each bit represents the projected occupation of points from immediately past the near map to the far map, as shown in **Fig. 11**. The geometry of such intermediate depth positions is stored in the occupancy frame. The EOM can be associated with both the 1-Map and 2-Map configurations. In the case of a 1-Map configuration, the Enhanced Occupancy Map (EOM) code can define up to 15 positions from the projected depth. In the

case of a 2-Map configuration, the number of EOM codes is limited to the surface thickness between the near and the far depth (e.g., a thickness of 4 is shown in **Fig. 11**). The attributes associated with those EOM points are stored in the attribute frame. As with Raw patches, EOM attribute patches can be stored in an SV stream.

The usage of the EOM tool is signaled with the flag `asps_eom_patch_enabled_flag`.

V-PCC Profiles and Conformance Points

Essentially, a profile of a codec corresponds to a selection of tools that can be used by an encoder and must be implemented by a conforming decoder. A profile is expected to represent a category of industrial applications of the codec. The selection of tools in a profile is a tradeoff between quality and bearable complexity for either the encoder or the decoder. This can be important where, for example, a low-end smartphone with capture and coding capabilities may not have the same coding or decoding performance as a high-end production workstation. Profiles can address diverse needs by emphasizing, for example, low-complexity decoding, low-latency distribution, or studio-grade quality.

A V-PCC profile is comprised of a codec group, a toolset profile component (directed to decoding the atlas), and a reconstruction toolset profile component (see annex A of Ref. 8 for a comprehensive description).

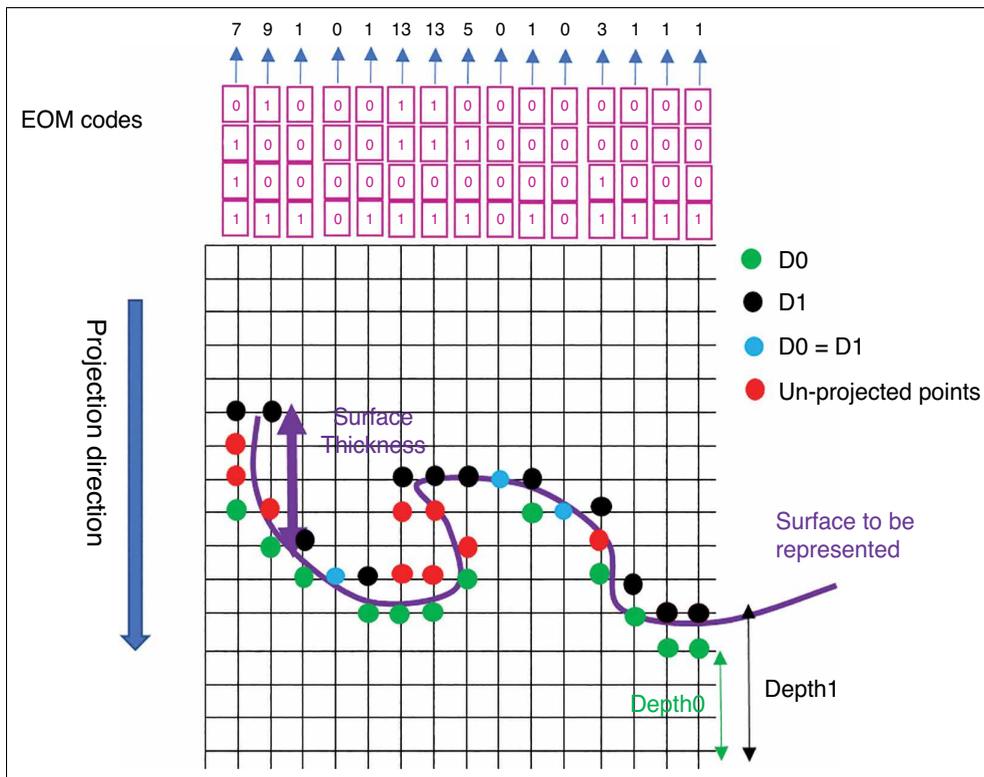


FIGURE 11. EOM, 2-Map case, the near map (green), the far map (black). If near and far map is the same point (blue), EOM code is set 0; otherwise, each bit of the EOM code corresponds to the consecutively further in-between points (in red).

During the development of the V-PCC profile requirements, the MPEG-3 DG group agreed that profiles need to cover at least the decoding of atlases with associated information but made the reconstruction process optional. Indeed, some companies were supportive of using proprietary tools for reconstruction to allow for differentiation and leveraging of algorithms or architecture of existing resources on the rendering device. Others wanted guarantees about the visual quality of the reconstructed point cloud. Thus, a peculiarity of V-PCC is that two conformance points are defined, which are illustrated in **Fig. 4**.

- The first, conformance point A, covers the decoded video subbitstreams and atlas metadata subbitstream. It also covers the patch sequence decompression. It does not, however, cover the 2D to 3D reconstruction process. Conformance point A is bit-accurate, that is, hard conformance.
- The second, conformance point B, covers the reconstruction process and is not bit-accurate: conformance point B is soft conformance.

In this paper, profiles covering conformance points A and B are considered (profile descriptions are given in Ref. 8) and especially those enlisted in **Table 1**.

In the current version of the V-PCC standard, two toolset components (basic and extended) and three main reconstruction toolset components (Rec0, Rec1, and Rec2) are defined. Readers should note that the numbers X suffixing the reconstruction profile “RecX” are merely enumerators and are not indicative of complexity, quality, or any such ordering.

The basic profile component essentially does not enable EOM, PLR, extended projection, or eight orientations. The extended profile component removes these restrictions.

The Rec0 reconstruction profile component was designed so that the normative but optional reconstruction tools are ignored during the reconstruction process (i.e., the reconstruction process is performed with tools outside the scope of the standard). Rec0 may be adapted to either low-end devices or ecosystems with devices using proprietary reconstruction tools.

Rec1 and Rec2 reconstruction profiles enable most of the standardized reconstruction tools, with the difference

TABLE 1. Naming convention of profiles. Acronyms are not part of the standard and are used in the section “Experimental Results.”

Acronyms	Profile names
B0	HEVC Main10 V-PCC <i>Basic Rec0</i>
B1	HEVC Main10 V-PCC <i>Basic Rec1</i>
B2	HEVC Main10 V-PCC <i>Basic Rec2</i>
E0	HEVC Main10 V-PCC <i>Extended Rec0</i>
E1	HEVC Main10 V-PCC <i>Extended Rec1</i>
E2	HEVC Main10 V-PCC <i>Extended Rec2</i>

being that Rec1 comprises geometry grid-based smoothing, while Rec2 comprises PBF (occupancy map pruning). Rec1 and Rec2 may be better suited to operators who prefer keeping control of the reconstruction visual quality or preserving artistic intent. Indeed, Rec1 and Rec2 tools output information computed from the original point cloud at the encoder side. It is noted that a decoder implementing Rec1 or Rec2 might not use signaled reconstruction tools.

Additionally, a fourth reconstruction toolset component (RecUnconstrained) allows any reconstruction tool and associated processing to be used, or not, for the reconstruction stage, whether or not it is signaled in the bitstream—user’s choice.

Experimental Results

In this section, we present some benchmark results for different coding tools. The tools are evaluated using a set of objective metrics described in Refs. 17 and 18, which include peak signal-to-noise-ratios (PSNRs) of a point-to-point error (D1 or point metrics) and a point-to-plane error (D2 or plane metrics) for geometry, as well as PSNRs of color and reflectance attributes. These distortion metrics compare the original data with the reconstructed data and provide numerical values.

Geometric Distortions

For D1 and D2, both mean square error (MSE) and PSNRs are reported. **Figure 12** shows how D1 and D2 are computed. For D1, the comparison is such that the MSE between the reconstructed point b_i and the closest corresponding point a_j in the reference point cloud is calculated. For D2, the MSE is calculated between the reconstructed point b_i and the surface plane in the given reference test data. The reference test data provides surface normal information to facilitate the computation of the surface planes, avoiding potential variations in the choice of which face is used to project each CC. In the case of dynamic content, which corresponds to test

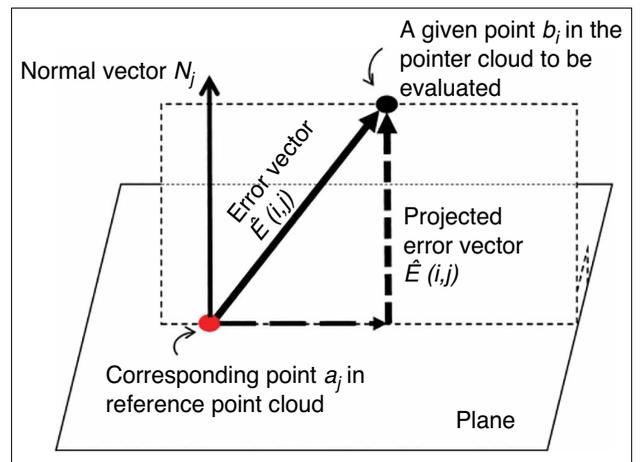


FIGURE 12. Point-to-point error E (D1) and point-to-plane error \hat{E} (D2) from Ref. 17.

TABLE 2. Naming convention of coding tools.

Acronyms	Coding tool	Default value
MC1/MC2	The number of maps, where MC1 means only one map, whereas MC2 means two maps	–
RAW	When mentioned, raw tool is activated. Required for lossless test	Off
EOM	When mentioned, enhanced occupancy map tool is activated	Off
SV	When mentioned, separate video tool for coding RAW and EOM geometry and attribute information for the atlas in a separate video. It is only applicable if RAW or EOM tools are activated	Off
GS0/GS1	Geometry smoothing (GS0 means deactivated, GS1 means activated). For lossless tests, GS is deactivated (GS0)	GS1
RDP0/RDP1	RDP0 means remove duplicate points tool is deactivated RDP1 means remove duplicate points tool is activated	RDP1
CGS	When mentioned, color grid smoothing is activated	Off
SPI	When mentioned, SPI is activated. When SPI is activated, MC1 is activated	Off
PBF	When mentioned, patch border filtering is activated. If PBF is activated, geometry smoothing is set to GS0	Off
PLR	When mentioned, point local reconstruction is activated. With PLR activated, MC1 is activated	Off
MS_AT11_AD11	Multiple streams tool is activated for coding geometry and attribute video streams. In this case, geometry and attribute of supplementary streams are coded as absolute values	Off
MS_AT10_AD10	Multiple streams tool is activated for coding geometry and texture video streams. In this case, geometry and texture of supplementary streams are coded as delta coding	Off

configurations discussed below, the reported distortion measures are averaged over all coded frames and further reported by type (i.e., average over I-frames, average over P-frames, and average over all frames).

Attributes Distortion

Color distortion is measured in “YUV” space as three separate MSE distortions, which are reported as PSNRs for each channel: Y, U, and V between the current point and its closest reference neighbor (D1).

Subjective Observations

Although subjective evaluation could not be conducted at the time of writing this manuscript because of coronavirus disease 2019 (COVID), some frame captures are provided to show the effects of the coding tools. To render point clouds for subjective evaluation, the reference point cloud renderer, chosen by MPEG 3DG group,¹⁹ is used. The “cube” rendering method is used here. This method was chosen by the MPEG community to best assess the impact of compression without introducing additional filtering at the rendering stage.

Evaluation Methodology

Naming Conventions

The tests are named in accordance with the following naming convention.

First, test names are prefixed B[0,2] or E[0,2], representing the acronyms from **Table 1**, thereby specifying

the targeted profile, as introduced in the section “V-PCC profiles and Conformance Points.”

Following the prefix, test names present a sequence of acronyms that identifies which coding tools are enabled. Note that when a tool is not mentioned in the test definition, a default value for the corresponding tool is introduced, as listed in **Table 2**.

Test Configurations

The software used to conduct the experiments is the test model developed within the MPEG community, called *TMC2 (Test Model for Category 2*, see Ref. 20). Revision 10.0 is available in Ref. 21 and the corresponding results are shown in Ref. 22.

Tests are run on each of seven 32-frame sequences, in both lossless and lossy configurations, to evaluate the performance of different coding tools using the objective metrics.

Tables 3 and **4** show, for these seven sequences, the summary of compression ratios and bitrates, respectively, for lossless and lossy configurations, relative to the MPEG “anchor” (i.e., reference) configuration (which is Basic.Rec1 with GS and RDP tools enabled; see Ref. 23). The right-hand portion of each table shows the average, minimum, and maximum bits per input point (bpip) of all sequences for the corresponding coding structure.

Table 4 reveals two key pieces of information: the first is the bitrates achieved by the V-PCC test model with respect to coding structure [all-intra (AI) and

TABLE 3. Compression ratio per sequence in percentage with respect to original uncompressed point clouds (i.e., basketball lossless AI represents 16% of original test model size). Followed by bits per input point (bpi) average, min and max among all sequences—lossless (E1_MC2_RAW_EOM) and lossy (B1_MC2_RDP1_GS1) 32 frames TMC2 R10.0. AI stands for all-intra, LD stands for low-delay, and RA stands for random-access.

	Basketball	Dancer	Longdress	Loot	Queen	Redandblack	Soldier	Average bpi	Min bpi	Max bpi
Lossless AI	16.34	16.66	27.76	16.51	15.13	23.76	18.55	10.67	8.17	14.99
Lossless LD	16.33	16.65	27.59	16.38	13.58	23.64	17.99	10.48	7.33	14.90
Lossy AI	0.34	0.39	1.46	0.59	0.63	0.85	0.92	0.41	0.12	0.92
Lossy RA	0.20	0.26	0.70	0.22	0.23	0.47	0.27	0.18	0.06	0.45

TABLE 4. Average lossy compression bitrate (Mb/s) on all sequences for each predefined rate (R01.5 from common test conditions; see Ref. 23) followed by bits per input point (bpi) average, min, and max among all rates—B1_MC2_RDP1_GS1 32 frames TMC2 R10.0. AI stands for all-intra and RA stands for random access.

	R01	R02	R03	R04	R05	Average bpi	Min bpi	Max bpi
Lossy AI	4.44	6.72	10.84	18.81	34.00	0.41	0.12	0.92
Lossy RA	2.38	3.24	4.80	8.30	17.34	0.18	0.06	0.45
Average	3.41	4.98	7.82	13.55	25.67	0.30	0.09	0.59

random-access (RA)]; the second is the average, minimum, and maximum bpi for all rates by the coding structure. The takeaway from **Table 4** is that bitrates achieved by V-PCC (using HEVC as an underlying video codec) are compatible with typical distribution networks deployed today.

Objective Performances

For the experiments presented in this section, four test families are identified: lossless, the lossy 1-Map tool (MC1), the lossy 2-Map tool (MC2), and the lossy 1-Map versus 2-Map tool. The last family preserves remarkable configurations from the two previous families.

For each test, a table presents the gain (negative value in green) or loss (positive value in red) as a percentage relative to the anchor (the first line on each table) for the different metrics.

For the lossless family, **Table 5** presents four data points concerning the bit per input point ratio (bpi) metrics: the whole bitstream (bpi total), geometry plus data (bpi data+geometry), geometry-only information (bpi geometry), and attributes-only information (bpi color).

For the three lossy families, the corresponding tables (**Tables 6–8**) present the ratios of the point metrics (point/D1), plane metrics (plane/D2), luminance (Y), and chrominance (U and V columns). These metrics are introduced in sections “Geometric Distortions” and “Attributes Distortion.”

For all of the families, the time ratios, in percentage gain, consumed inside the encoder (self-time enc) and decoder (self-time dec) are shown, as well as the memory consumption in each of the encoder (gain mem enc) and decoder (gain mem dec).

TABLE 5. Metric and performance ratios of lossless tools compared to the anchor B1_MC2_RAW (in percentage).

Experiments	Bpi total	Bpi data +geometry	Bpi geometry	Bpi color	Self-time enc	Self-time dec	Gain mem enc	Gain mem dec
B1_MC2_RAW	0	0	0	0	0	0	0	0
B1_MC2_RAW_SV	-0.08	0.29	0.31	-0.15	-2.1	-0.46	-1.83	-3.38
E1_MC2_RAW_EOM	-2.71	-18.05	-20.08	-0.44	-3.42	-7.23	1.22	-1.01
E1_MC2_RAW_EOM_SV	-2.76	-17.93	-19.95	-0.51	-5.11	-8.28	0.71	-2.13
E1_MC1_RAW_EOM	1.34	-3.81	-99.55	2.18	-6.06	-21.49	-8.09	-13.44

TABLE 6. Metric and performance ratios of 1-Map tools compared to the anchor B0_MC1_GS0_RDP0 (in percentage).

Experiments	Point/D1	Plane/D2	Y	U	V	Self-time enc	Self-time dec	Gain mem enc	Gain mem dec
B0_MC1_RDP0_GS0	0	0	0	0	0	0	0	0	0
B1_MC1_RDP0_GS1	-1.14	-4.81	0.89	0.34	0.68	0.33	73.02	0	0.61
B1_MC1_SPI_RDP1_GS0	-36.02	5.19	-0.74	3.36	2.73	40.93	66.38	37.86	55.96
B1_MC1_SPI_RDP1_GS1	-47.6	-21.3	-0.22	3.78	3.33	45.91	177.73	37.84	56.77
B2_MC1_RDP0_PBF	2.71	-3.02	0.61	-0.04	0	0.28	-3.24	2.5	-6.18
E1_MC1_PLR_RDP1_GS0	-34.65	3.16	2.33	7.05	6.13	46.91	67.99	37.87	55.97
E1_MC1_PLR_RDP1_GS1	-47.14	-21.08	2.89	7.42	6.81	47.18	175.93	37.86	56.79
E2_MC1_PLR_RDP1_PBF	-46.49	-23.24	2.67	6.86	5.91	45.44	62.02	38.55	45.17

TABLE 7. Metric and performance ratios of 2-Map tools compared to the BO_MC2_RDP0_GS0 (in percentage).

Experiments	Point/D1	Plane/D2	Y	U	V	Self-time enc	Self-time dec	Gain mem enc	Gain mem dec
B0_MC2_RDP0_GS0	0	0	0	0	0	0	0	0	0
B1_MC2_RDP1_GS0	0.01	-0.04	-0.32	0.13	0.16	-0.4	-5.76	-7.5	-14.06
B1_MC2_RDP0_GS1	-19.85	-24.77	1.09	0.3	0.54	0.21	89.03	-0.02	0.72
B1_MC2_RDP1_GS1	-19.36	-23.56	0.92	0.5	0.78	-0.19	55.24	-7.51	-13.66
B1_MC2_RDP1_GS1_CGS	-19.36	-23.55	92.93	126	139.56	0.16	76.39	-6.61	-7.66
B1_MC2_RDP1_GS1_MS_AT11_AD11	30.83	28.17	64.29	72.33	71.89	79.27	69.02	-6.22	-10.83
B1_MC2_RDP1_GS1_MS_AT10_AD10	30.06	24.62	12.1	15.61	14.91	66.44	77.15	-0.01	-4.93
B2_MC2_RDP0_PBF	-18.79	-27.4	0.46	-0.21	-0.26	-1.99	-4.94	1.24	-5.8
B2_MC2_RDP1_PBF	-18.71	-27.37	0.04	-0.05	-0.12	-0.89	-7.78	-6.36	-19.98

TABLE 8. Metric and performance ratios of 1-Map versus 2-Map tools compared to the B0_MC1_GS0_RDP0 (in percentage).

Experiments	Point/D1	Plane/D2	Y	U	V	Self-time enc	Self-time dec	Gain mem enc	Gain mem dec
B0_MC1_RDP0_GS0	0	0	0	0	0	0	0	0	0
B0_MC2_RDP0_GS0	-36.42	3.69	0.27	5.95	4.9	53.88	74.13	36.16	90.75
B1_MC1_SPI_RDP1_GS1	-47.6	-21.3	-0.22	3.78	3.33	45.91	177.73	37.84	56.77
B1_MC2_RDP1_GS1	-46.31	-22.02	0.43	6.44	5.64	53.59	170.32	25.93	64.7
B2_MC2_RDP1_PBF	-44.67	-24.98	0.26	5.88	4.73	52.51	60.58	27.5	52.63
E1_MC1_PLR_RDP1_GS1	-47.14	-21.08	2.89	7.42	6.81	47.18	175.93	37.86	56.79
E2_MC1_PLR_RDP1_PBF	-46.49	-23.24	2.67	6.86	5.91	45.44	62.02	38.55	45.17

A discussion follows each test series. In the last section, some screen captures are presented, highlighting some visual effects corresponding to the coding tools.

Lossless Results

Here the anchor is the B1_MC2_RAW, which is a 2-Map configuration with the Raw tool activated.

The lossless results in **Table 5** show that the EOM tool (E1_MC2_RAW_EOM) performs better with a gain of 2.71% for the whole bitstream (bpip total). The benefit occurs essentially because of the geometry coding gain of 20.08% (bpip geometry). This tool is less time-consuming, showing a gain of 7% on the decoder

decoding time (self-time dec) and has no significant impact on memory consumption.

When combined with the SV tool, EOM is the tool that performed the best of all tested lossless configurations in terms of performance (E1_MC2_RAW_EOM_SV).

Lossy 1-Map Results

Here, the anchor is the B0_MC1_RDP0_GS0, which is a 1-Map configuration with no other tool.

SPI (B1_MC1_SPI_RDP1_GS0) and PLR (E1_MC1_PLR_RDP1_GS0) are tools that enhance the reconstructed point cloud with a remarkable gain of 35% in the point metric at a cost of around 70% in complexity

on the decoder side. However, SPI and PLR show a loss in the Y, U, V metrics, with the PLR tool showing a little bit greater loss.

Geometry results are even better when map-enhancement and filtering tools are combined (B1_MC1_SPI_RDP1_GS1, E1_MC1_PLR_RDP1_GS1, and E2_MC1_PLR_RDP1_PBF). GS and PBF improve point metrics by about 12% and plane metrics by about 15%. They have the strong benefit of reducing aliasing effects along patch borders. Compared with the anchor, one advantage of the PBF tool is that it adds neither complexity (self-time enc/dec) nor memory consumption overhead (gain mem enc/dec) when combined with PLR (E2_MC1_PLR_RDP1_PBF versus E2_MC1_PLR_RDP1_GS0).

Lossy 2-Map Results

Here the anchor is the B0_MC2_RDP0_GS0, which is a 2-Map configuration with no enhancement tool.

As seen from **Table 7**, the Remove Duplicate Point (RDP) tool does not bring significant improvement to the metrics on its own, but when combined with other tools (e.g., GS: B1_MC2_RDP0_GS1 versus B1_MC2_RDP1_GS1), it decreases the decoding complexity by about 34% (self-time dec), as fewer points are handled downstream in the reconstruction process.

One remarkable tool is the PBF (B2_MC2_RDP0_PBF), which performs better than the anchor in terms of all metrics and complexity. This tool yields a 18.79% gain in the point metric, 27.4% in plane metric, and though holding roughly the same in Y, U, V, it has a gain of 2% on the encoder (self-time enc), 5% on the decoder (self-time dec), and 6% for decoder memory usage (gain mem dec). Here again, combining the PBF and the RDP tools reduces the complexity and memory consumption on both the encoder and decoder sides.

Lossy 1-Map Versus 2-Map Results

To be able to compare the 1-Map versus 2-Map tool, the best previous tests from the respective families are collected and listed in **Table 8**.

Here, the anchor is the B0_MC1_RDP0_GS0, which is a 1-Map configuration with no enhancement tool.

Comparison of these two families leads to several observations. First, all selected tools bring a gain in the geometric metrics (point and plane) with a compromise on the color (around 5%). This net gain in quality comes at the cost of an increase in complexity for both the encoder and decoder (self-time enc, self-time dec), including their memory consumption (gain mem enc, gain mem dec).

The PBF tool (B2_MC2_RDP1_PBF) brings a 44.67% gain in the point metric and a 24.98% gain for the plane, at an increase of 60% in complexity (self-time enc and self-time dec) compared to the anchor. Coupled with the PLR (E2_PLR_RDP1_PBF), the PBF tool sees PLR's advantages, keeping a low complexity, and benefits



FIGURE 13. Longdress reference source model frame #1051 overview and closeups.

from the visual improvement as described in the following section.

Some Visual Comparisons on Lossy Configurations

In this section, we present some screen captures to highlight various tool benefits. Screen captures are taken from the MPEG 3DG test content called *longdress*, at the middle rate R03 targeted in the common test conditions (see Ref. 23). The original uncompressed test point cloud (reference source model) was provided by 8i (see Ref. 24) and is shown in **Fig. 13**.

Comparisons Regarding the Patch-Filtering Processes

Here, using 2-Map configurations, the comparison is among the tools that impact patch border alignment, namely GS and PBF. The reference source model and 2-Map without smoothing are presented as the basis for comparison. As seen in **Fig. 14**, filtering the patch border, with GS1 or PBF, significantly increases the final point cloud quality as it strongly reduces patch border aliasing (see shoulder areas). The benefit is two-fold. First, the geometry artifacts are reduced, leading to more accurate silhouettes and surfaces. Second, the photometric continuity is better preserved at patch boundaries.

Comparisons Regarding the Number of Maps

For this evaluation, the comparison is between tools that change the number of maps, say B0_MC1_RDP0_GS0

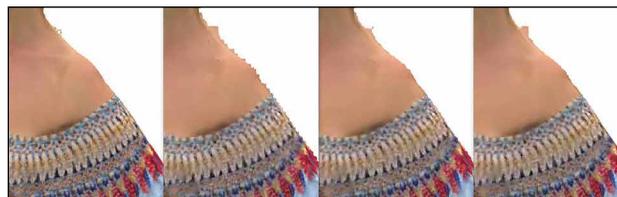


FIGURE 14. Smoothing tools. From left to right: reference source, no smoothing (B0_MC2_RDP0_GS0), GS on (B1_MC2_RDP1_GS1), PBF on (B2_MC2_RDP1_PBF).

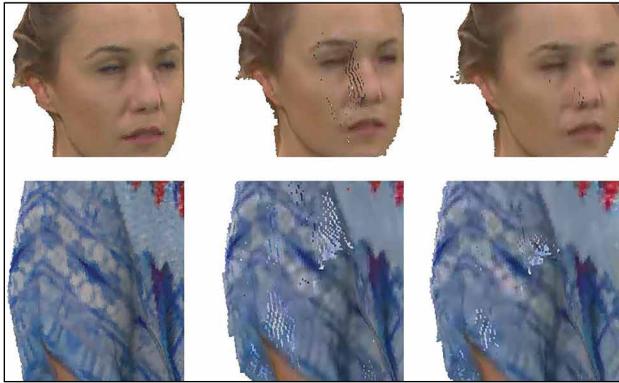


FIGURE 15. Number of maps tools. From left to right: reference source, 1-Map (B0_MC1_RDP0_GS0) and 2-Map (B0_MC2_RDP0_GS0).

for 1-Map and B0_MC2_RDP0_GS0 for 2-Map. The results, as shown in **Fig. 15**, demonstrate that increasing the number of projected points, thanks to the use of several maps, visually enhances the reconstructed point cloud. A 2-Map configuration captures more accurately the distribution of points in 3D space; fewer holes appear; and surface continuities are better preserved, even if there are still artifacts due to the lossy nature of the compression [see the section “Multiple Maps (Named MC for Map Count)”].

Comparisons Regarding the Map-Enhancement Tool

In **Fig. 16**, using the 1-Map tool, the comparison is among the map-enhancement tools: SPI and PLR. The reference and unfiltered 1-Map are given as the basis for comparison. As can be seen, artificially increasing the number of points at the point cloud reconstruction stage by using SPI or PLR leads to significant visual enhancements, especially at the surface continuity level. In this set of experiments, PLR generally led to the highest quality visual results, at the cost of a marginal increase in data transmitted, though still showing artifacts due to lossy compression.

Comparisons Regarding the Combination of Tools

Finally, **Fig. 17** illustrates the benefits to the reconstruction of combining the tools together. While some tools

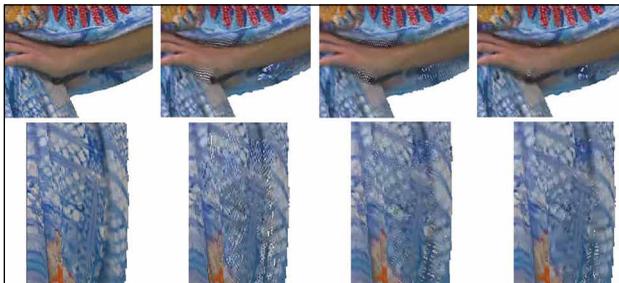


FIGURE 16. Map-enhancement tools. From left to right: reference source, 1-Map no enhancement (B1_MC1_RDP0_GS0), 1-Map and SPI active (B1_MC1_SPI_RDP1_GS0), 1-Map and PLR active (E2_MC1_PLR_RDP1_GS0).

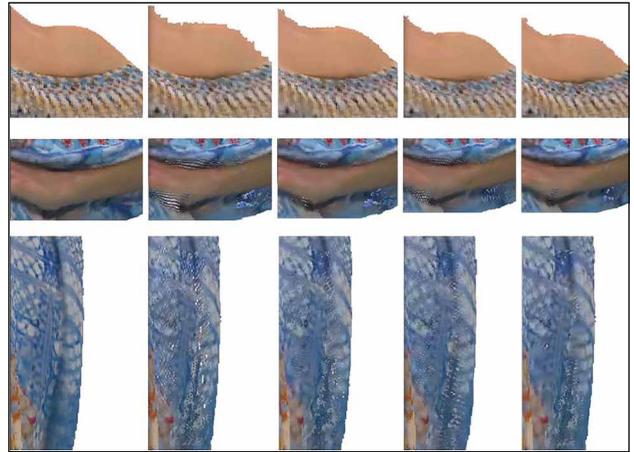


FIGURE 17. Combination of tools. From left to right: reference source, 1-Map no enhancement (B0_MC1_RDP0_GS0), 2-Map with GS (B1_MC2_RDP1_GS1), 1-Map with SPI and GS (B1_MC1_SPI_RDP1_GS1), and 1-Map with PLR and PBF (E2_MC1_PLR_RDP1_PBF).

focus on PBF and trying to reduce the aliasing effect on geometry and photometry, other tools contribute to the advantage of filling holes introduced by compression, whether resulting from unprojected points or 2D compression errors. In most cases shown, one can see that superior visual results are achieved by using 2-Map or 1-Map using PLR with PBF. The choice of which combination to use would thus be made according to the targeted device performance (see the section “Lossy 1-Map Versus 2-Map Results”).

Conclusion

The V-PCC standard is the first MPEG codec for point cloud compression. Its publication by ISO as International Standard is planned in Q2/2021. V-PCC was initiated with near-term deployment in mind; therefore, its architecture was designed to rely upon conventional 2D video codecs. It thus leverages the large base of video codecs already deployed in consumer electronics devices [e.g., advanced video coding (AVC), HEVC, etc.] and, at the same time, it is expected to directly scale up in performance, thanks to coming generations of video codecs (e.g., VVC).

This document has presented the main tools implemented in the V-PCC standard and their performance (i.e., quality for a given bitrate at an acceptable complexity for mass market devices) in the V-PCC MPEG test model. Although performance may differ for an optimized product implementation, according to the targeted platform architecture and capabilities, the V-PCC test model is the platform used by the MPEG group to make the decisions to include tools in the V-PCC standard, based on a balance between complexity and signal distortion. The tools evaluated enable reconstruction that maintains fidelity to the original point cloud or that provides bitrate reduction compatible with typical user network capabilities, at a

few megabits per second. It is interesting to note that according to the performance results, the extended profile tools may not require additional computational resources when compared to the basic profile tools and can even significantly reduce decoder complexity for some implementations.

V-PCC realtime decoding and rendering in an AR environment has already been demonstrated on mid-range mobile phones in trade fairs such as IBC2019 (see Refs. 25 and 26). Futurewei is releasing an open-source version of a V-PCC decoder, OpenV3C (see Ref. 27). In parallel, MPEG has worked on a convergence between the V-PCC and Metadata for Immersive Video (MIV; see MPEG-I part 12) standards, to provide a partly common specification, named “V3C” for “Visual Volumetric Video-based Coding,” to facilitate understanding and better promote the adoption of these 3D immersive standards by industry. Even now, V-PCC performance allows contemplating deployment in mass market applications as diverse as AR telepresence, VR for medical applications, free viewpoint video, edge cloud 3D video gaming, edutainment, remote learning, and many more immersive, innovative, creative services to be unleashed.

References

1. K. Cao, Y. Xu, Y. Lu, and Z. Wen, “Owlii Dynamic Human Mesh Sequence Dataset,” *ISO/IEC JTC1/SC29/WG11 m42816, 122th MPEG Meeting*, San Diego, CA, USA, Apr. 2018.
2. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “Use Cases for Point Cloud Compression,” Genova, Switzerland, w16331, Jun. 2016.
3. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “PCC Requirements,” Gwangju, South Korea, w17353, Jan. 2018.
4. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “PCC Requirements Status Update,” Alpbach, Austria, m53595, Apr. 2020.
5. High Efficiency Video Coding, Feb. 1, 2018. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.168+SCM-8.7
6. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “V-PCC Codec Description,” Alpbach, Austria, w19332, Apr. 2020.
7. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29/WG11, “PCC TMC2 With Additional Projection Plane,” Ljubljana, Slovenia, m43494, Jul. 2018.
8. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “Container for V-PCC Editing,” Alpbach, Austria, m53943, Apr. 2020. [Online]. Available: http://wg11.sc29.org/doc_end_user/documents/130_Alpbach/comments/comment-566-V-PCC_DIS_d90.zip
9. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “[PCC] TMC2 Pixel Interleaving in Geometry and Texture Layers,” Ljubljana, Slovenia, m43723, Jul. 2018.
10. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “PCC TMC2 Geometry Coding Improvements,” Gwangju, South Korea, m42111, Jan. 2018.

11. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29/WG11, “PCC TMC2 Low Complexity Geometry Smoothing,” Ljubljana, Slovenia, m43501, Jul. 2018.
12. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29/WG11, “[V-PCC] On Attribute Smoothing,” Geneva, Switzerland, m51003, 2019, Oct. 2019.
13. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29/WG11, “[V-PCC][specification] Patch Border Filtering Specification,” Geneva, Switzerland, m51501, Oct. 2019.
14. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29/WG11, “[VPCC-TM] New Contribution on Duplicate Point Avoidance in TMC2,” Macao, China, m44784, Oct. 2018.
15. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29/WG11, “[V-PCC] Report on Core Experiment CE 2.18 on Lossy Additional Points Patch for Improved Visual Quality,” Marrakech, Morocco, m45946, Jan. 2017.
16. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29/WG11, “[V-PCC][New Proposal] Generalized Enhanced Occupancy Map for Depth,” Geneva, Switzerland, m47895, Oct. 2019.
17. D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, “Geometric Distortion Metrics for Point Cloud Compression,” *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, pp. 3460–3464, 2017, doi: 10.1109/ICIP.2017.8296925.
18. R. Mekuria, S. Laserre, and C. Tulvan, “Performance Assessment of Point Cloud Compression,” *IEEE Visual Comm. Image Process. (VCIP)*, St. Petersburg, FL, pp. 1–4, 2017, doi: 10.1109/VCIP.2017.8305132.
19. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “[VPCC][software] mpeg-pcc-renderer Software v5.0,” Alpbach, Austria, m53592, Apr. 2020.
20. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “V-PCC Test Model v10,” Alpbach, Austria, w19325, Apr. 2020.
21. Git Reference for Test Model v10. Jun. 3, 2020. [Online]. Available: <http://mpegx.int-evry.fr/software/MPEG/PCC/TM/mpeg-pcc-tmc2/commit/1018e5467ea0d18b879272bcb8c060e1cf497a9>
22. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “V-PCC Performance Evaluation and Anchor Results,” Alpbach, Austria, w19327, Apr. 2020.
23. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “Common Test Conditions for PCC,” Alpbach, Austria, w19324, Apr. 2020.
24. E. d’Eon, B. Harrison, T. Myers, and P. A. Chou, “8i Voxelized Full Bodies, Version 2—A Voxelized Point Cloud Dataset,” *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) Input Document m40059/M74006*, Geneva, Switzerland, Jan. 2017.
25. C. Guede, R. Schaefer, and IBC 2019, Sep. 2019. [Online]. Available: <https://www.interdigital.com/videos/v--pcc-the-first-mpeg-codec-for-point-cloud-compression>
26. S. Schwarz, M. Pesonen, and IBC 2019, “Real-Time Decoding and AR Playback of the Emerging MPEG Video-Based Point Cloud Compression Standard,” Sep. 2019.
27. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) JTC1/SC29 WG11, “OpenV3C—Multiplatform Open Source Implementation of V-PCC,” Alpbach, Austria, w19375, Apr. 2020.

About the Authors



Céline Guede received an engineering degree from the Polytech Orleans School, Orleans, France, in 1998, and immediately joined Technicolor as an R&D engineer. After gaining specialization in software development, she joined the Advanced Television Systems Committee (ATSC) 3.0 standardization group on next-generation broadcast in 2013, where she contributed to the evaluation and selection of scalable high efficiency video coding (SHVC)/HEVC video encoding and Moving Picture Experts Group (MPEG) MMT transport technologies. She shifted her focus to point cloud compression at the end of 2016, actively contributing several technology proposals to the emerging MPEG Point Cloud Compression standard [Video-based Point Cloud Compression (V-PCC), known as *MPEG-I Part 5*]. Since 2019, she has been an architect at InterDigital where she continues to actively contribute on the V-PCC standard evolution. She is currently working on the dissemination of the V-PCC technology through realtime codec demonstrations and the adoption in future TV broadcast standards such as Sistema Brasileiro de Televisão Digital (SBTVD).



Pierre Andrivon received an advanced master's degree in signal and image processing from Télécom Paris, Paris, France, in 2006. As a research engineer at Vitec, Paris, France, he developed H.264/AVC encoders and contributed to collaborative projects on realtime video processing. In 2011, he joined Technicolor, where he successively participated in the standardization of H.265/HEVC and SHVC video codecs, and the single layer high dynamic range (SL-HDR) technology jointly developed with Philips. He was a co-editor of the European Telecommunications Standards Institute Technical Specification (ETSI TS) 103 433 multipart standard. Since 2019, as a principal scientist for InterDigital, he leads the effort on point cloud and mesh compression. He contributed to the standardization of the MPEG V-PCC standard as a co-editor of the specification. He has been involved in several video standardization committees: MPEG, digital video broadcasting (DVB), ETSI, Blu-ray Disc Association (BDA), SMPTE, high definition multimedia interface (HDMI), and Consumer Technology Association (CTA). He filed more than 100 patent applications and holds +20 issued patents, mainly on video processing, video coding, and color management.



Jean-Eudes Marvie is a principal scientist at InterDigital. He received a PhD degree from the University of Rennes, Rennes, France, in 2004. His research interests are realtime rendering, large model's visualization and generation, procedural modeling and rendering, and mixed reality.

He previously applied these techniques to the field of live broadcast, interactive previsualization for cinema production, and immersive experiences in home ecosystems such as telepresence.



Julien Ricard is an architect on the point cloud compression team at InterDigital, Rennes, France. He participated in the development of the response to the MPEG call for proposals on the compression of point clouds and his response took second place on nine answers. Since then, he has

co-authored several important proposals accepted in the MPEG V-PCC standard. For the past three years, he has been the MPEG-3 DG-PCC software coordinator for the candidate video point cloud compression test model and the official MPEG-3 DG point cloud rendering engine. In this position, he coordinates the integration of new proposals from MPEG promoters into the test model and guarantees the conformity and the quality of the source codes. To promote and evangelize around the point cloud format and the compression of this data, he participated in the creation of several demonstrations on mobile devices to show the ability to render, decode, and stream this type of data in realtime. He holds a doctorate in computer science and a master's degree in computer engineering from the University of Lyon, Lyon, France.



Bill Redmann's career has always mixed technology and entertainment. While earning his master's degree in engineering from the University of California, Los Angeles (UCLA), he built practical props for *Battlestar Galactica* and *Buck Rogers* at Universal Studios, Los Angeles. As the director

of technology for Disney Imagineering and Virtual World Entertainment, he researched, developed, and fielded simulator ride systems and virtual reality attractions. At Technicolor, Rennes, France, an interest in standards in the context of digital cinema led him to join SMPTE, where he eventually became a SMPTE Fellow. Now as an InterDigital's Director of Standards for Immersive

Media Technologies, he contributes to next-generation television systems. He holds 71 issued U.S. patents relating to digital cinema, virtual and augmented reality, media production, distributed network streaming, interactive systems, displays, online communities, mobile content distribution, exercise tracking, travel planning, and electric vehicle infrastructure. His two favorites—a roller coaster you design then ride, and a keyboard for dolphins to communicate with people—both really worked.



Jean-Claude Chevet joined Thomson, Rennes, France, in 1983 as an electronic technician at the TV R&D plant, after graduating from the Institut Universitaire de Technologie of Angers, France, in electrical engineering domain. He worked on micro-controller programming for advanced devices development in consumer electronic domain. In 1987, he worked on home automation systems, then on video processing for the HD TV and plasma TVs. In 1997, he received an engineer diploma from Institut des Techniques d'Ingénieur de l'Industrie, Auxerre, France, after

three years of education at Ecole Nationale Supérieure de Physique de Strasbourg (ENSPS). He then worked as a research engineer in the video compression lab at an MPEG-4 player development company in Rennes, France, and then in the network area (5 GHz, 802.11a). Chevet also worked on embedded systems for an H264 professional encoding equipment for the Grass-Valley company (formerly Harmonics), Rennes, France. He also worked in the human perception area (visual attention model and reframing) to develop demonstrators based on this technology and then joined the image editing and enhancement area (image processing). In 2015, he joined the ATSC 3.0 project at Technicolor as a developer of technical demonstrators integrating SHVC/HEVC video encoding. Since 2017, he has worked in the field of compression of volumetric video, notably he participated in several contributions for the new MPEG standard, in particular, for the 3DG Ad hoc Group on V-PCC.

*Presented at the SMPTE 2020 Virtual Conference Event, 10–12 November 2020.
Copyright © 2021 by SMPTE.*



Take the Lead

SMPTE Corporate Memberships lift your organization to the next level

Are you taking full advantage of all SMPTE can offer? We've introduced new Corporate Memberships levels designed for every type of business, whether you're an industry leader, newly minted start-up, small business or consultant. With professional and executive education, conferences and events – plus a full array of opportunities for corporate promotion, standards development and student support – you'll spark success for your whole team.

Review the benefits of corporate membership, and lock in a lower rate by renewing now, before prices increase in July.

**Renew before July
and save!**



Visit [smpte.org/corporate-membership](https://www.smpte.org/corporate-membership)