Point Cloud Coding: The *Status Quo*

João Ascenso, Fernando Pereira
Instituto Superior Técnico – Universidade de Lisboa
Instituto de Telecomunicações
Lisboa – PORTUGAL

IEEE International Conference on Multimedia and Expo (ICME)
London, UK
10th July 2020

Tutorial Outline

1. 3D Visual Representation and Coding (F.Pereira)
2. Plenoptic Function based Imaging (F.Pereira)

Questions

3. Point Cloud Coding: Basic Approaches (F.Pereira)

Questions

4. Point Cloud Coding: Standardization, part 1 (J.Ascenso)

Questions

Point Cloud Coding: Standardization, part 2 (J.Ascenso)

5. Point Cloud Quality Assessment (J.Ascenso)
6. Summary and Trends (J.Ascenso)

Questions

TÉCNICO LISBOA

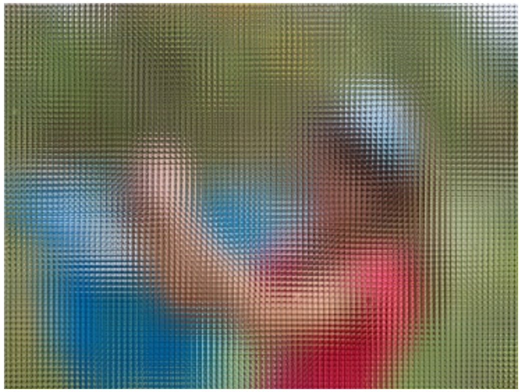
4 Point Cloud Coding: Standardization

TÉCNICO LISBOA

4.1 JPEG Pleno

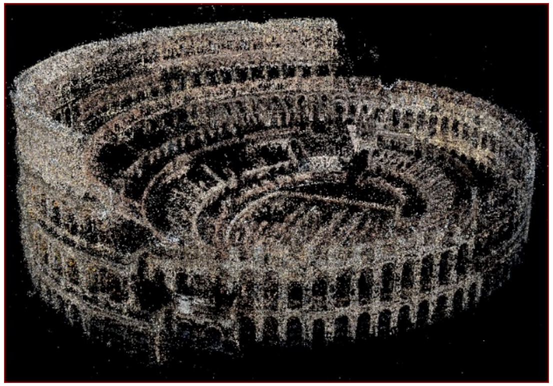
JPEG PLENO: Coding for Emerging Imaging Modalities

JPEG PLENO targets a standard framework for the representation and exchange of new imaging modalities such as light field, point cloud and holographic imaging

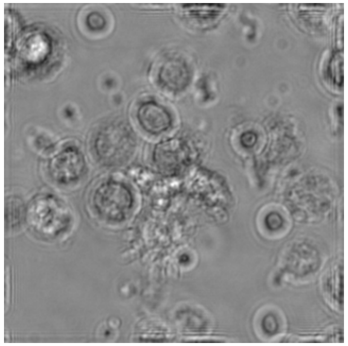


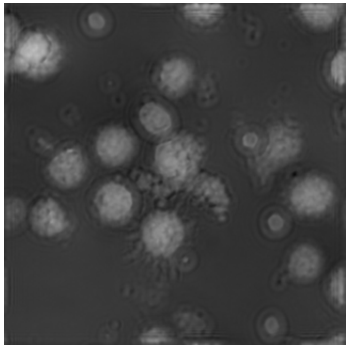
Lenslet light field

High density camera array



Point cloud





Microscopic holography

JPEG PLENO: Building a Framework

- Representation models
- Coding solutions
- Metadata
- File format
- Systems tools

JPEG PLENO: Achievements and Ongoing Work

Light fields

- Specification of light fields coding standard for L**enslet** and **High Density Camera Array (HDCA) light fields**

Point clouds

- Study and definition of objective quality metrics and subjective quality assessment protocols
- Definition of point cloud coding requirements and common test conditions
- **Call for Evidence** with focus on coding technologies for static point cloud content that enable scalable decoding of the bitstream and random access.

Holography

- Collection of test data, definition of numerical reconstruction software
- Definition of holographic coding requirements and use cases as well as common test conditions
- **Call for Proposals** to be issued soon

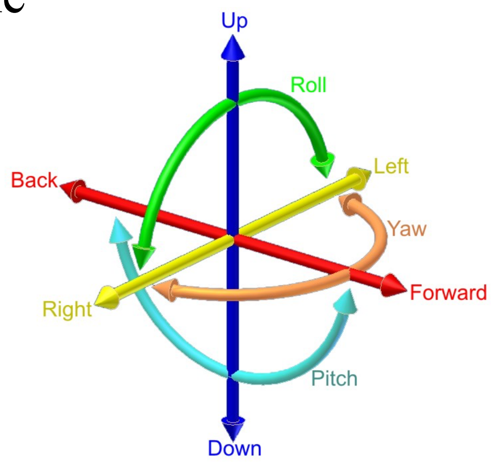
TÉCNICO LISBOA

4.2

MPEG Point Cloud Coding Standards

MPEG-I: Coding for Immersive Applications

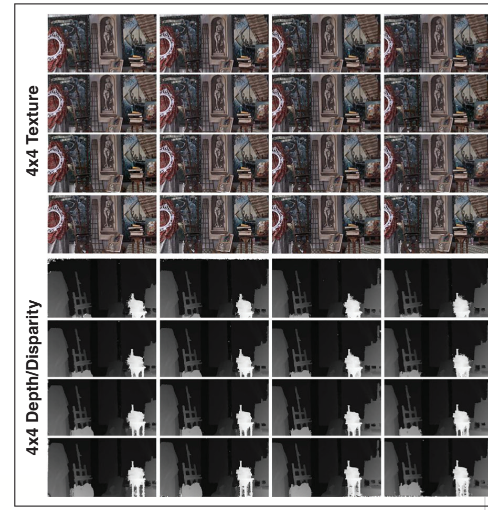
- Enables visualization of content from several viewpoints, thus supporting free navigation around regions of interest in the scene
  - 3DoF: 3 degrees of freedom
  - 3DoF+: adds motion parallax with limited range
  - 6DoF: large ranges of freedom of movement

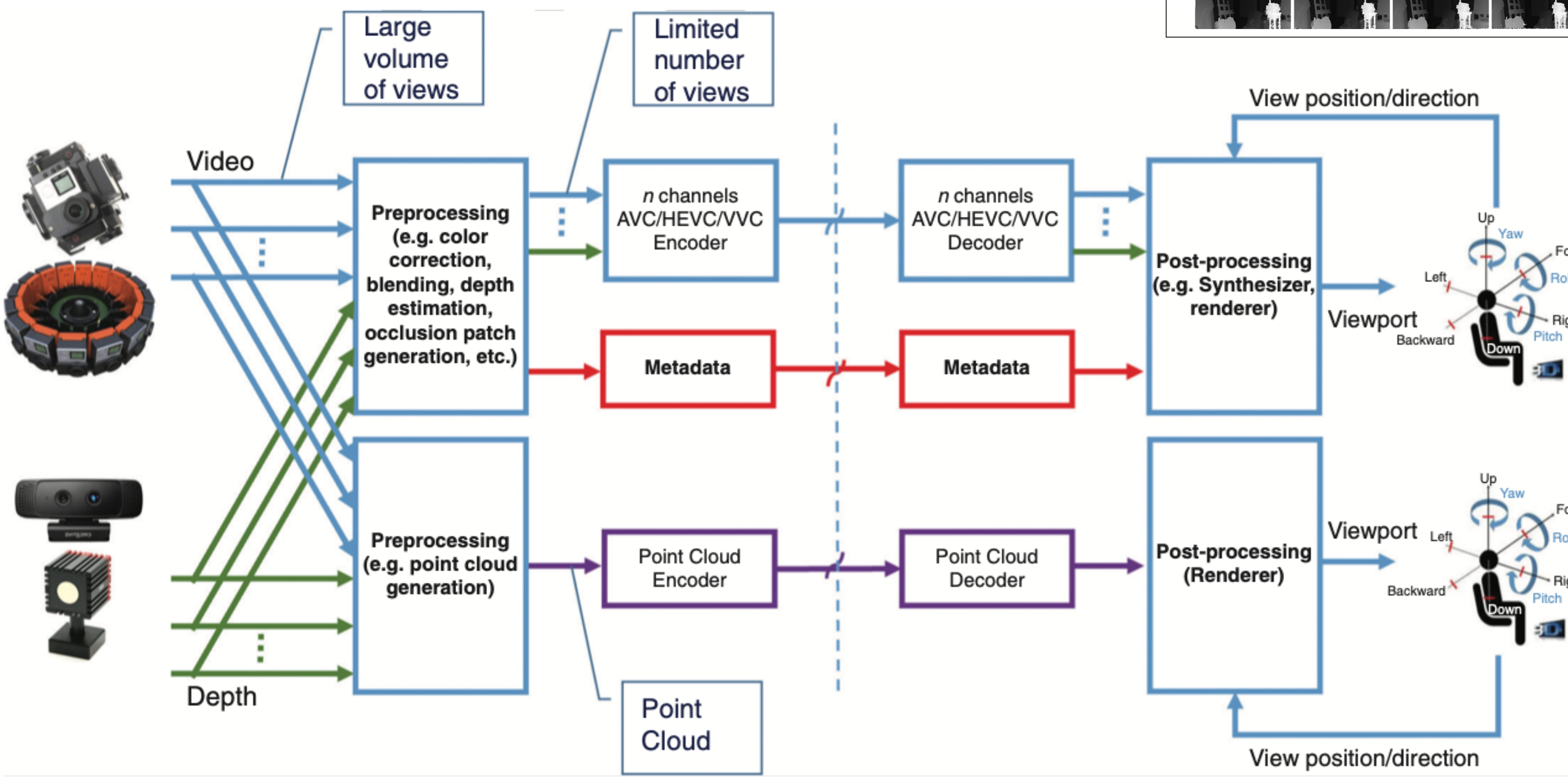


- MPEG-I targets:
  - Head mounted VR and AR devices
  - Light-field displays

- MPEG-I coding technologies are under exploration:
  - MPEG Video Group: standard 2D video coding + depth + metadata
  - MPEG 3DG group: point cloud coding technologies

MPEG-I Processing and Coding Pipeline





From G. Lafruit, D. Bonatto, C. Tulvan, M. Preda, and L. Yu, "Understanding MPEG-I Coding Standardization in Immersive VR/AR Applications", SMPTE Periodical, December 2019.

TÉCNICO LISBOA

Immersive Video

Courtesy: Technicolor/InterDigital.

16–25 camera feeds in UHD@30fps lead to 150–240 Mbits/sec

From G. Lafruit, D. Bonatto, C. Tulvan, M. Preda, and L. Yu, “Understanding MPEG-I Coding Standardization in Immersive VR/AR Applications”, SMPTE Periodical, December 2019.

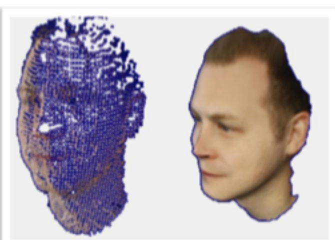
MPEG Point Cloud Compression

V-PCC 01/2020

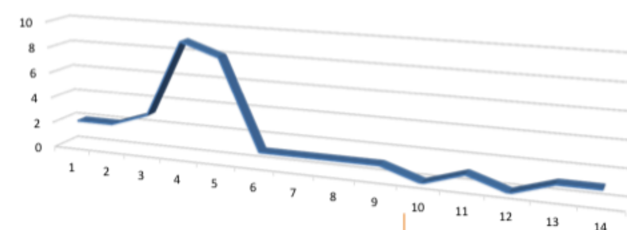
G-PCC 4/2020

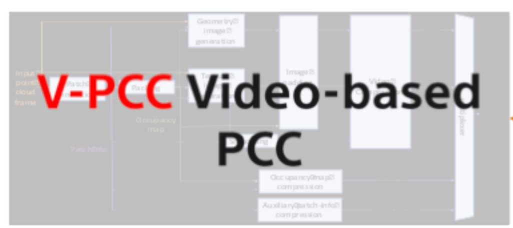
2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020

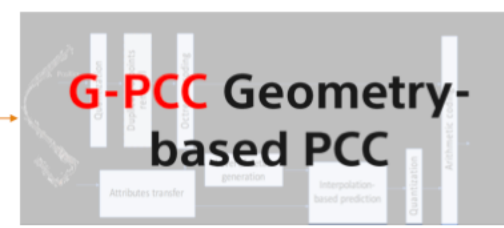
MPEG initiated the work on PCC → **In April 2017 MPEG issued a Call for Proposals** → **First Committee Draft issued in October 2018**



9 technology leading companies responded and MPEG evaluated them in October 2017



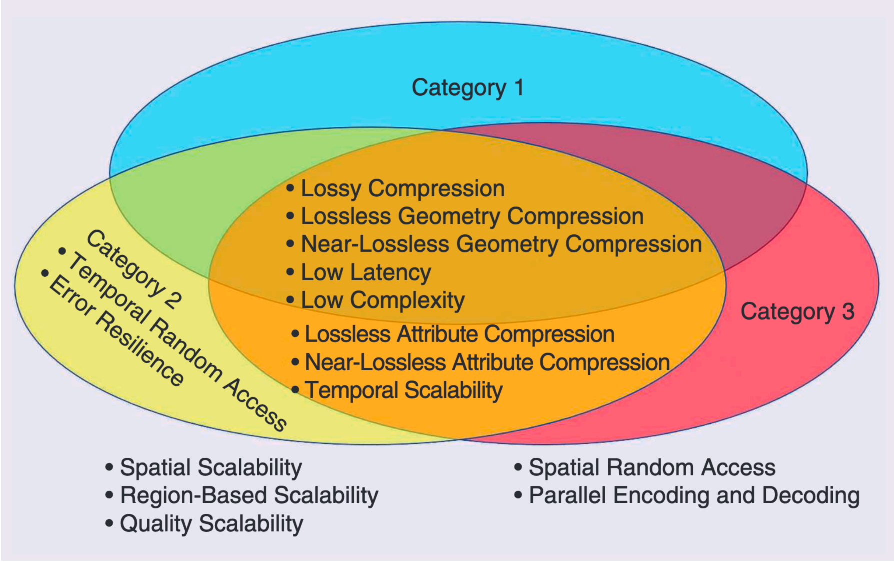




from Danillo Graziosi

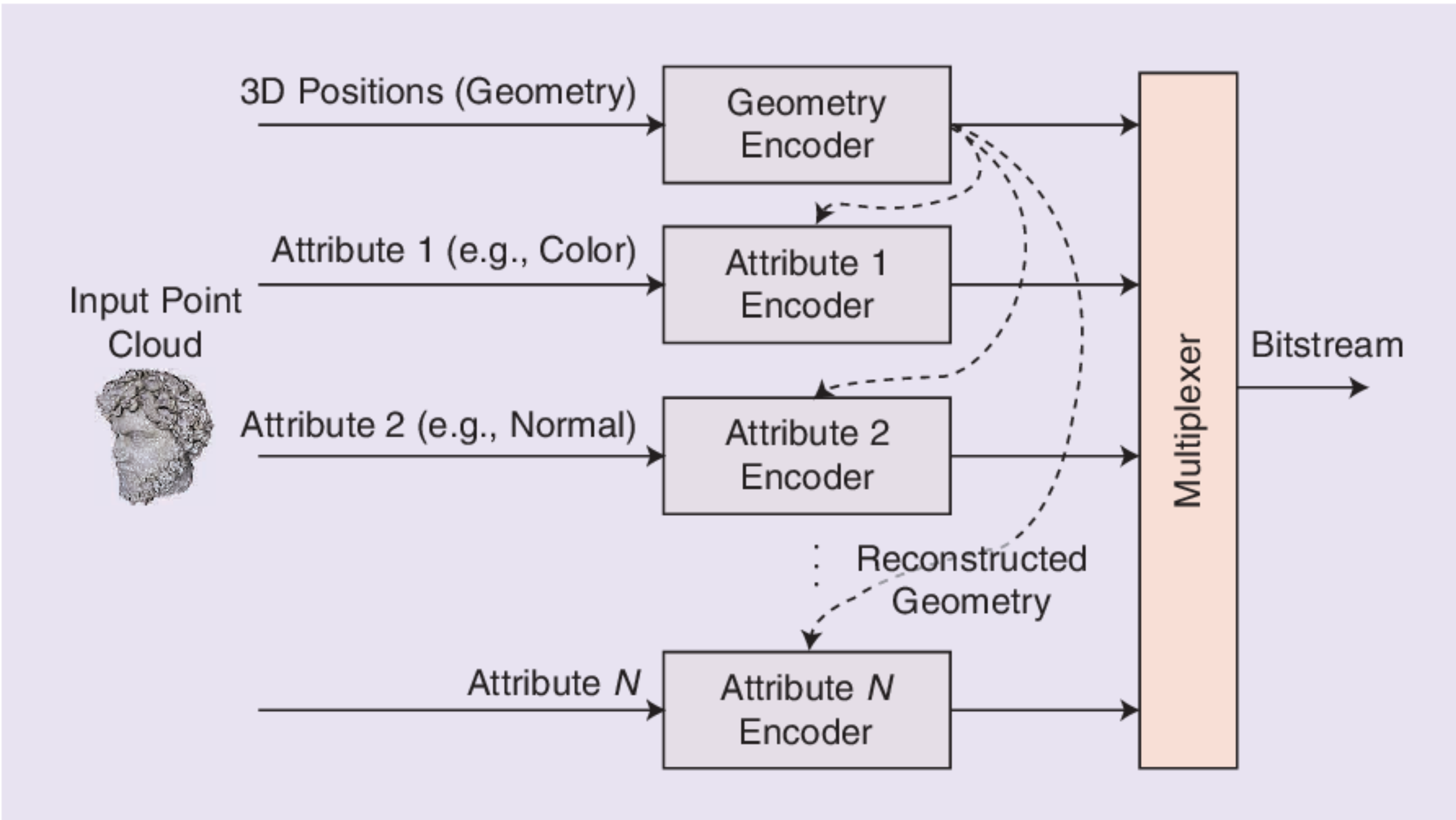
MPEG PCC Organization

- MPEG organizes the point cloud coding landscape into three categories:
  - Category 1: static point clouds for GIS and cultural heritage objects and collections
  - Category 2: dynamic point clouds for VR/AR and tele-immersive applications
  - Category 3: dynamically acquired or fused point clouds, such as autonomous navigation based and large-scale 3D dynamic maps
- Categories 1 and 3: **MPEG G-PCC** geometry compression
- Category 2: **MPEG V-PCC** video compression approach



from L.Cui et al. "Point-cloud compression: Moving Picture Experts Group's new standard in 2020", IEEE Consumer Electronics Magazine, Jan. 2019.

Many Codecs: MPEG Conceptual View

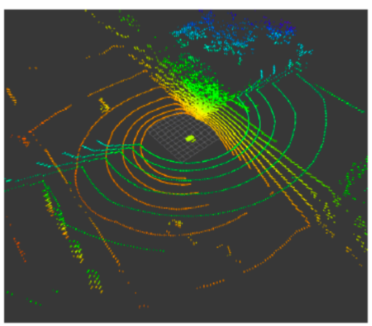


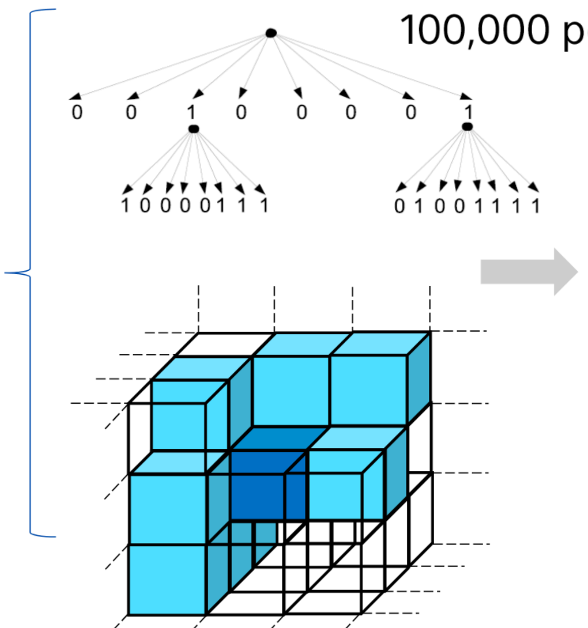
from L.Cui et al. "Point-cloud compression: Moving Picture Experts Group's new standard in 2020", IEEE Consumer Electronics Magazine, Jan. 2019.

Point Cloud Coding: MPEG G-PCC

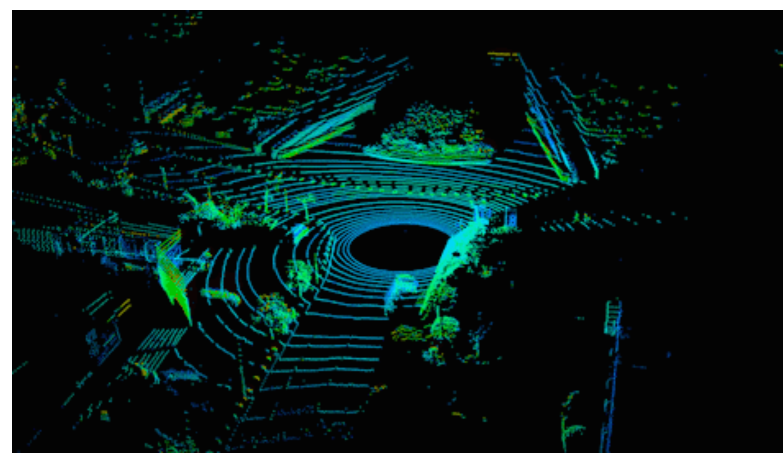
G-PCC (ISO/IEC 23090-9) Overview

- ★ **MPEG Geometry-based Point Cloud Compression (G-PCC):**
  - Going for Draft International Standard (DIS) in April 2020
  - Suitable for static and progressive PC coding
- ★ Capable of **lossy** and **lossless coding** of large point clouds
  - Combination of octree and surface reconstruction technologies
  - Supports random access and parallelization





100,000 points @ 10 fps → 110 Mbps (uncompressed)

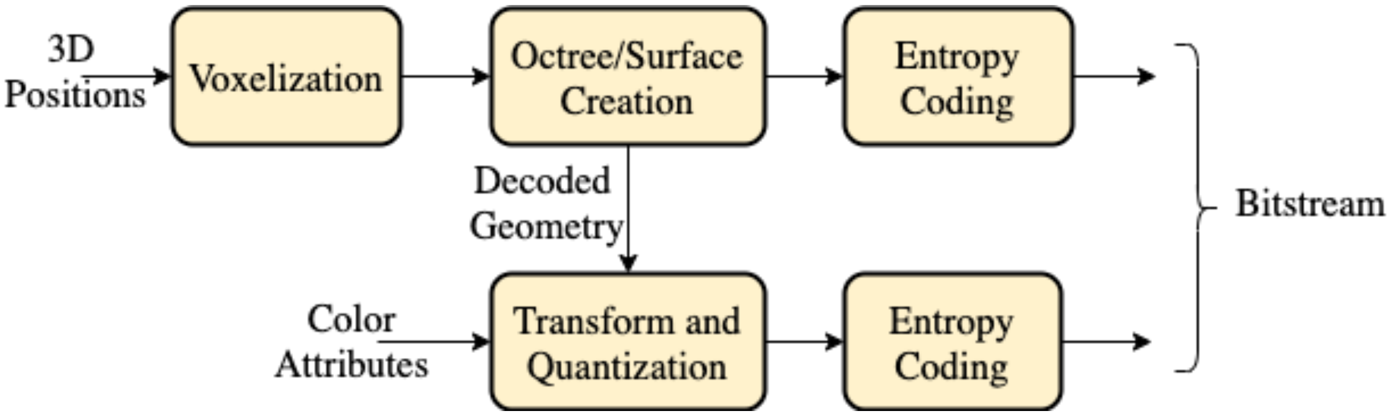


24 Mbps (lossless)

from Danillo Graziosi

G-PCC Principles and Foundations

- All encoding processing is made with integer values and thus voxelization is applied first
- Geometry coding is based on **octree** and **surface** models for point cloud data structuring
- Attributes are dependent upon the encoded geometry and a **recoloring** process has to map one or more original color values to decoded points
- Color coding uses a **region-adaptive hierarchical transform** or an **interpolation-based prediction** to exploit the **spatial correlation**

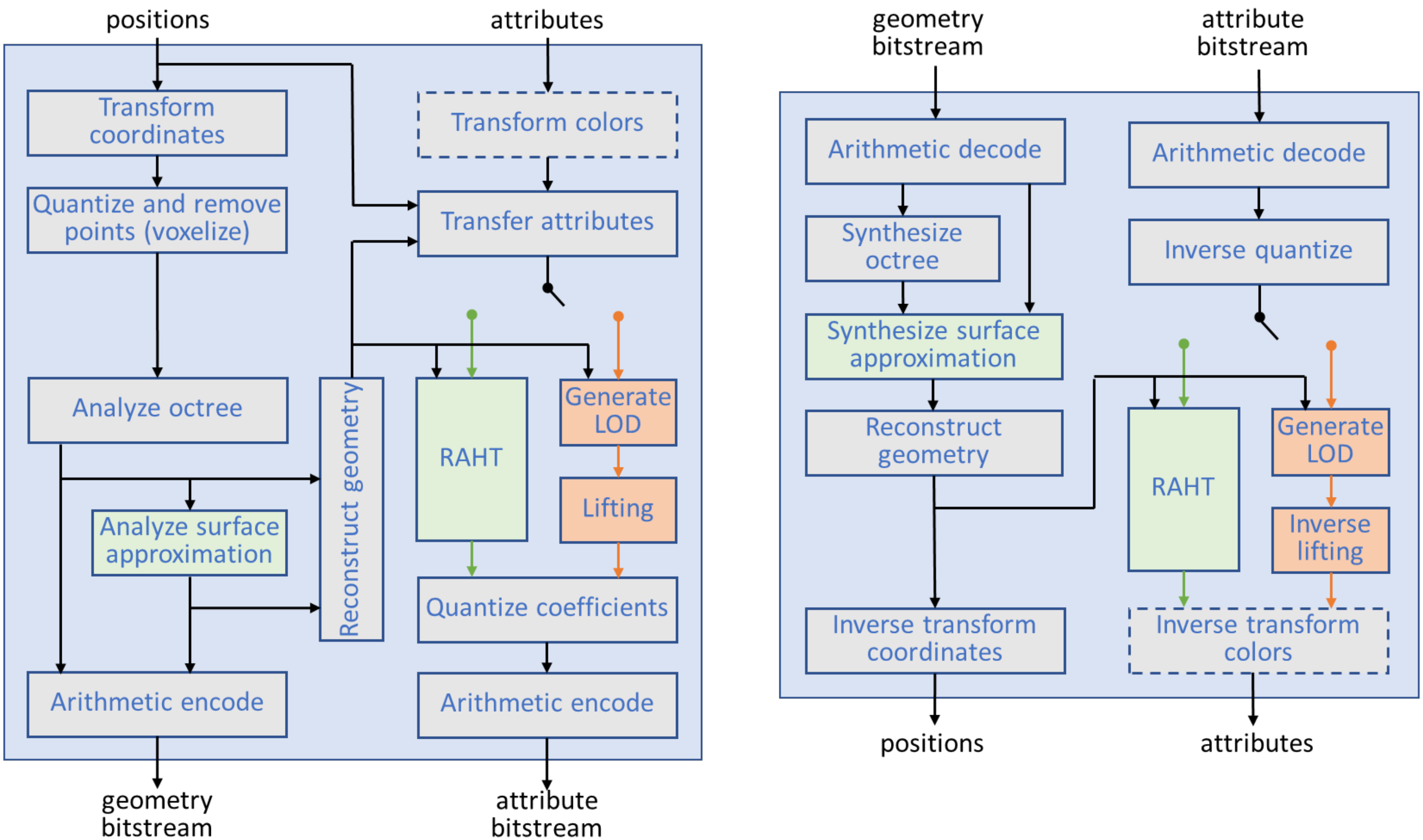


G-PCC Geometry Coding

- **Point cloud data** is organized into:
  - Slices: set of points independent decodable (cannot overlap)
  - Tiles: groups of slices defined by a bounding box (can overlap)
- Two possible geometry coding modes:
  - **Lossless mode:** Blocks have size of 1×1×1 (include only one voxel)
  - **Lossy mode:** Blocks are 2×2×2 or larger (contain more than one voxel)
- Two types of geometry coding approaches:
  - Octree: geometry is represented as an octree from the root all the way down to a leaf level of individual voxels (suitable for category 3 data)
  - Trisoup: octree decomposition from the root down to some leaf level and a surface model that approximates the point cloud surface within each leaf (suitable for category 1 data)

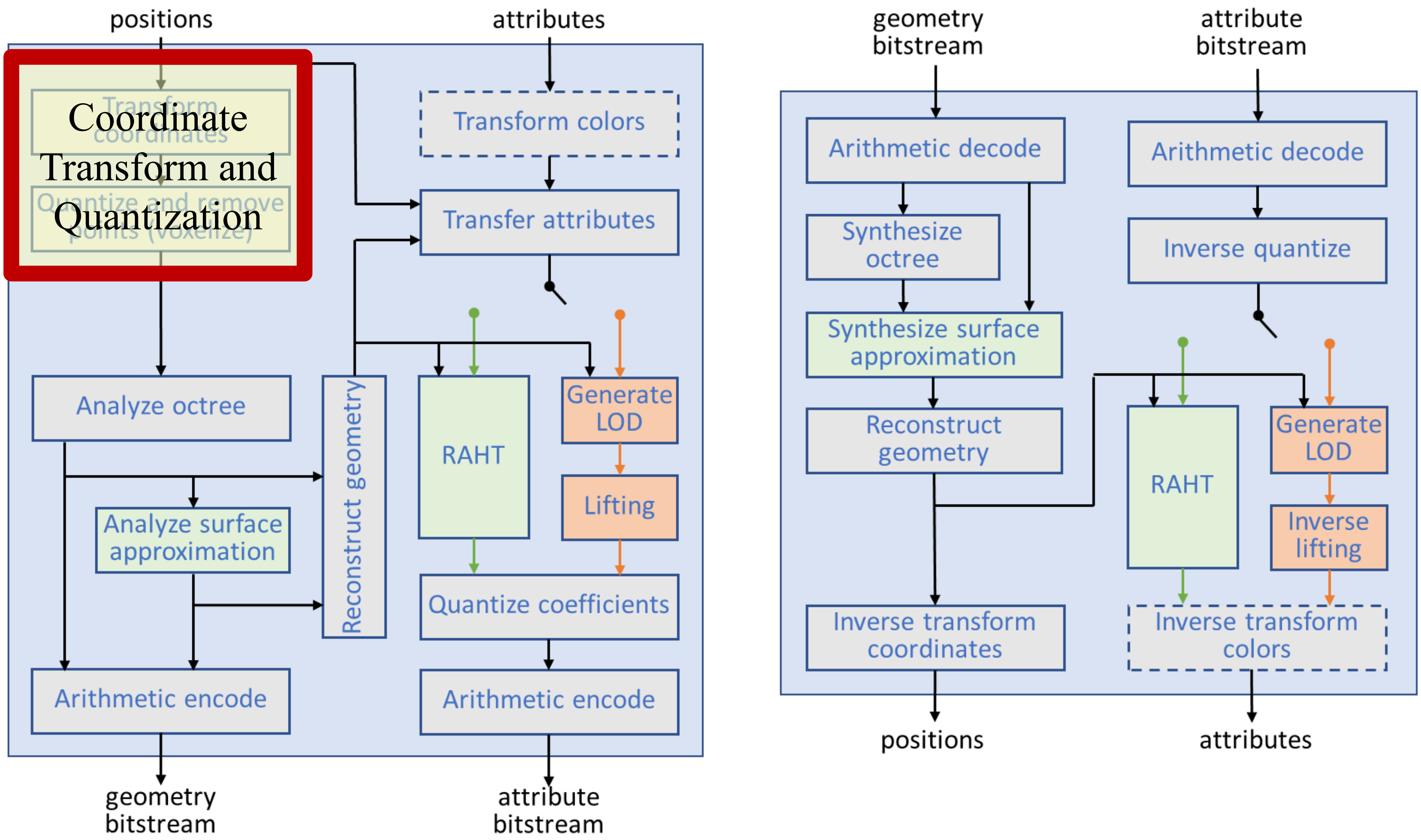
G-PCC Architecture

Green modules are typically used for Category 1 data. Orange modules are typically used for Category 3 data. Other modules are common between Categories 1 and 3.



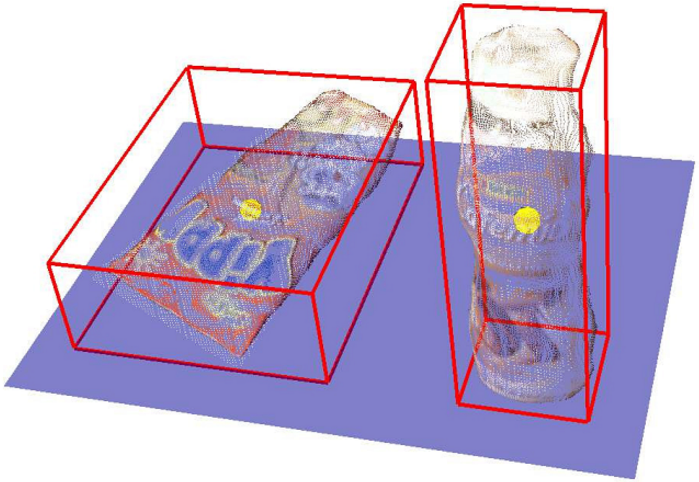
From K. Mammou et al., "PCCTest Model Category 13 v3," ISO/IEC JTC1/SC29/WG11 N17762, Ljubljana, Slovenia, Jul. 2018.

G-PCC Architecture



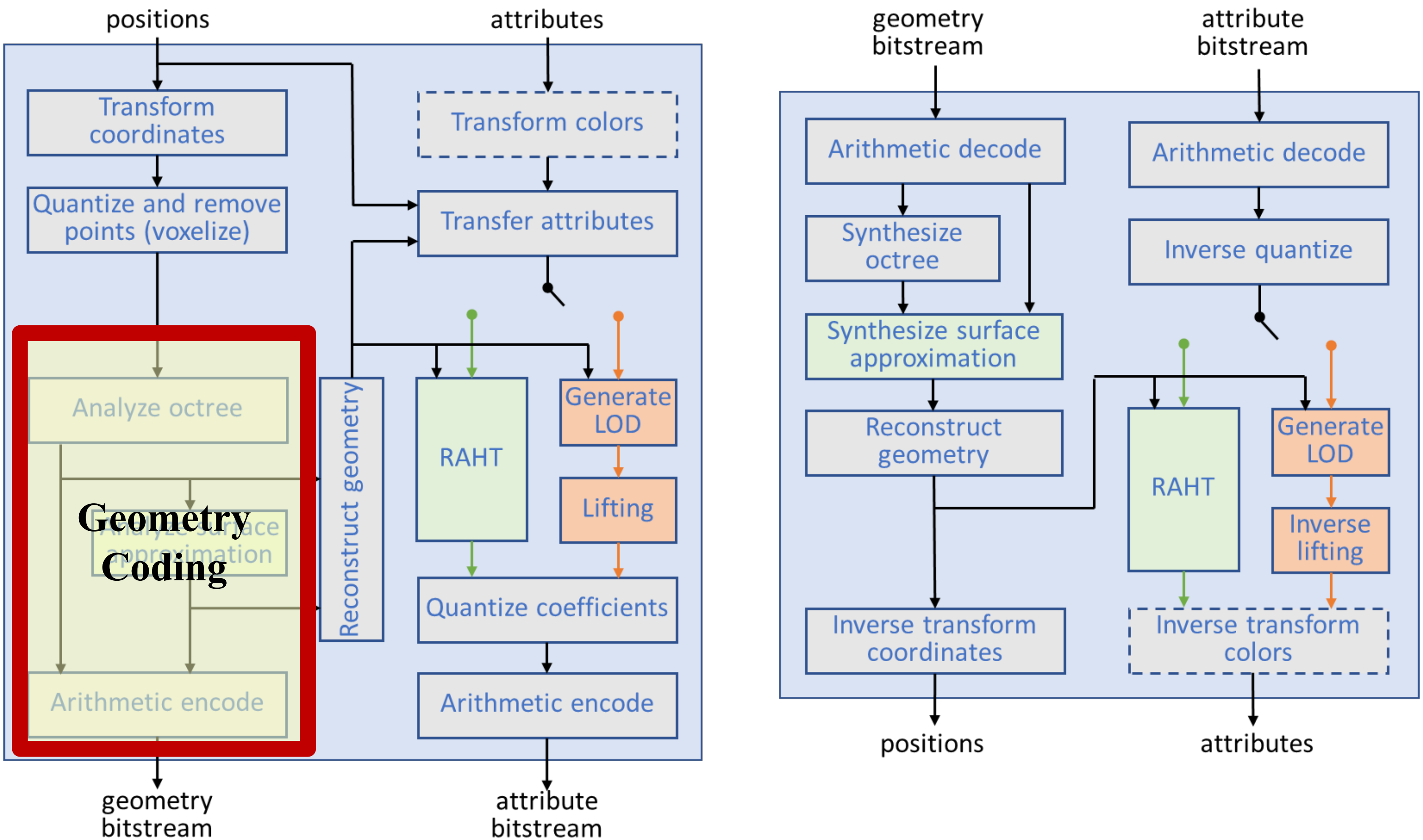
G-PCC Coordinate Transform and Quantization

- 3D world coordinates are transformed with **scaling and translation operations**
  - Resulting x, y, z values lie in the cube $[0,2^d - 1]^3$, where d is the octree target depth
  - Parameters for these transformations are sent to the decoder
- Point positions are **quantized** (i.e. rounded) and associated to voxels with non-negative d-bit integer coordinates
  - Duplicate points are removed



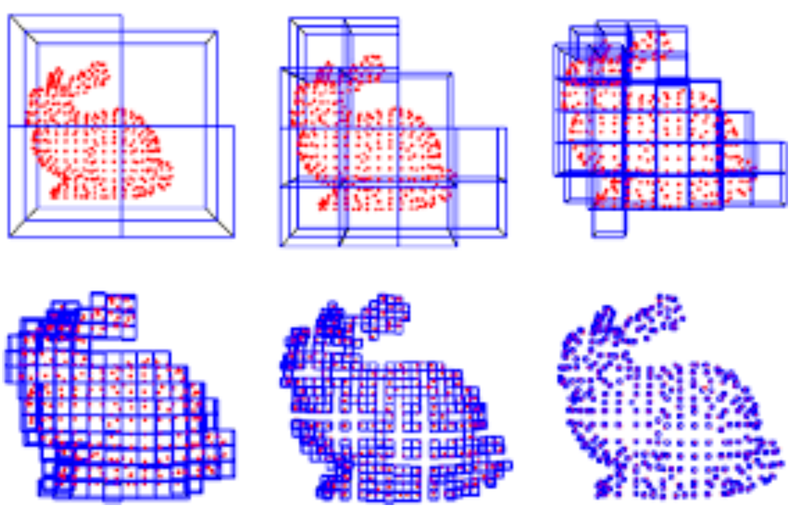
M. Richtsfeld and M. Vincze, "Robotic Grasping of Unknown Objects", Computer Science, 2011

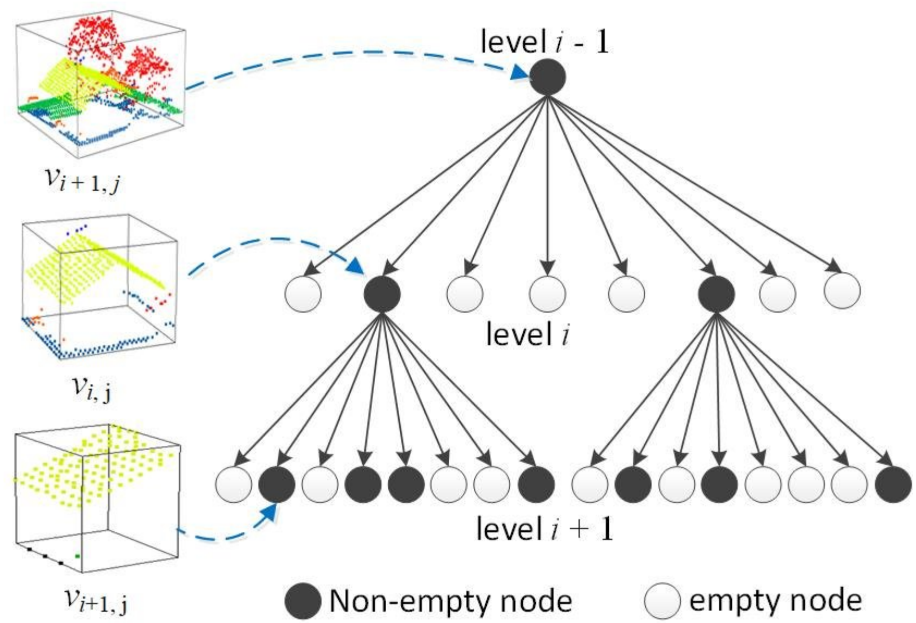
G-PCC Architecture



Geometry Coding: Octree Model

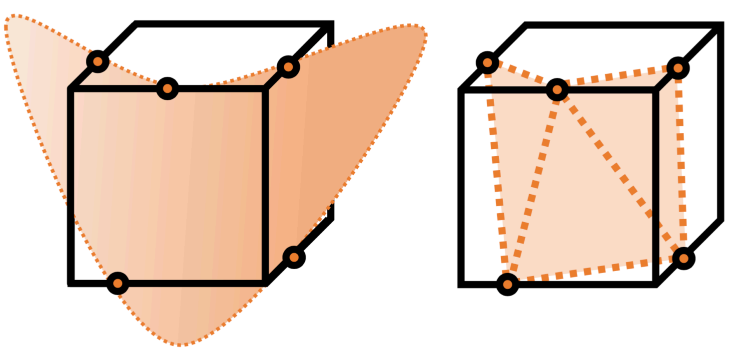
- The entire 3D volume is partitioned into blocks of voxels with a fixed width
- Block-level geometry is then represented as an octree with a target depth
  - Blocks correspond to the octree leaf nodes
- **Occupancy of the blocks** is encoded with one byte per node of the octree
- **Binary arithmetic coding of the occupancy bytes** is performed according to a **probability model** that depends on the occupancy states of the already-coded neighboring blocks



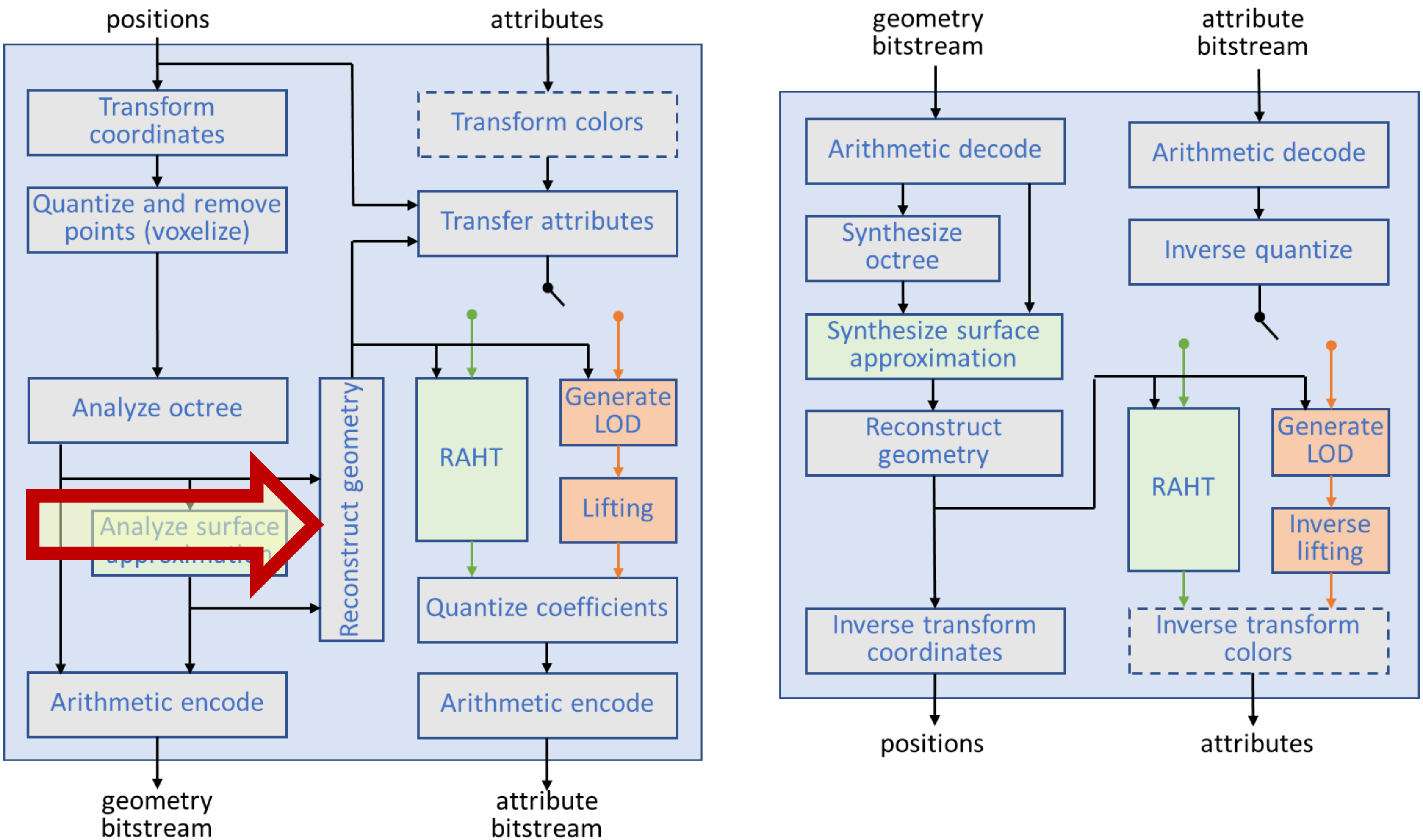


Geometry Coding: Surface Model

- Surface model used is a triangulation with 1-10 triangles per block, resulting in a triangle soup (known as "Trisoup" geometry codec)
- **For the lossy mode,** geometry within each block is represented as a surface intersecting each block edge at most once
  - Represent the points inside a block as a collection of triangles and thus, rather complex shapes (i.e. non-planar polygons) can be modelled
  - Bit vector determines which edges contain a vertex (i.e. intersection with the block edges) and which do not
  - For each edge containing a vertex, the position of the vertex along the edge is uniformly scalar quantized
  - Bit vector and quantized vertex positions are arithmetically coded



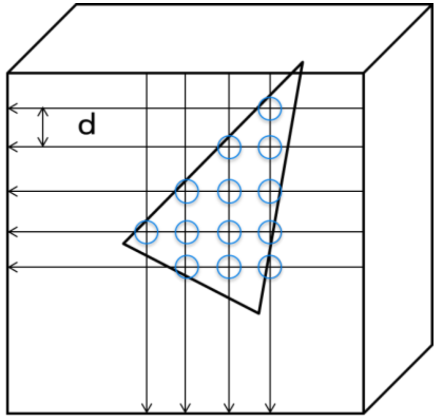
G-PCC Architecture



Geometry Reconstruction

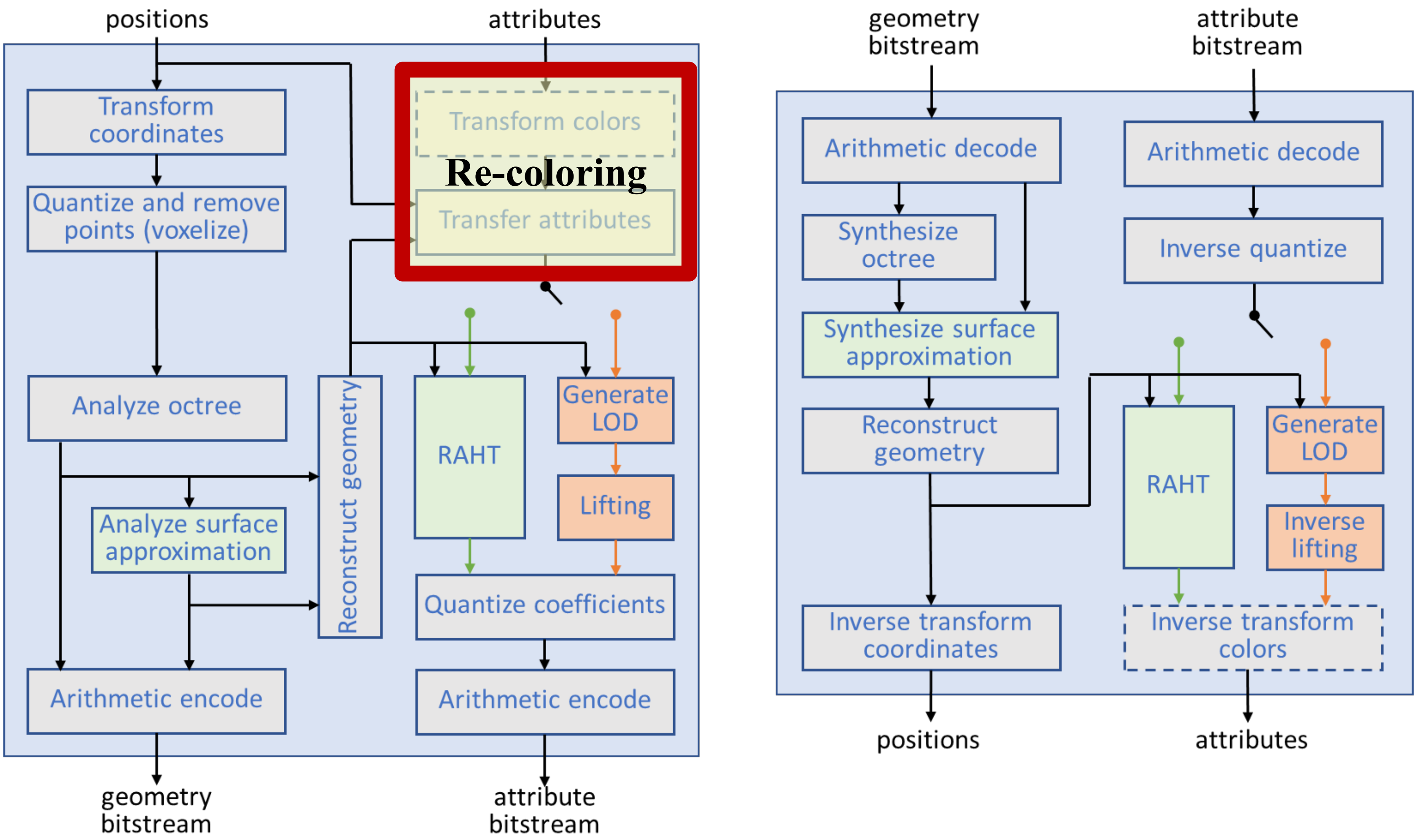
- **Blocks and vertices decoding:**
  - Octree part of the geometry is reconstructed
  - Vertices are decoded for each block and the triangulated surface is reconstructed

- **Surface reconstruction:** decoded geometry corresponds to the regularly-spaced points on the surface of the triangles
  - Aims to create geometry at a spatial resolution greater than or equal to the spatial resolution of the color information
  - In practice, a point is created for each voxel position intersecting the triangle surface



from 3DGroup, "G-PCC codec description v3" N18486, April 2018.

G-PCC Architecture

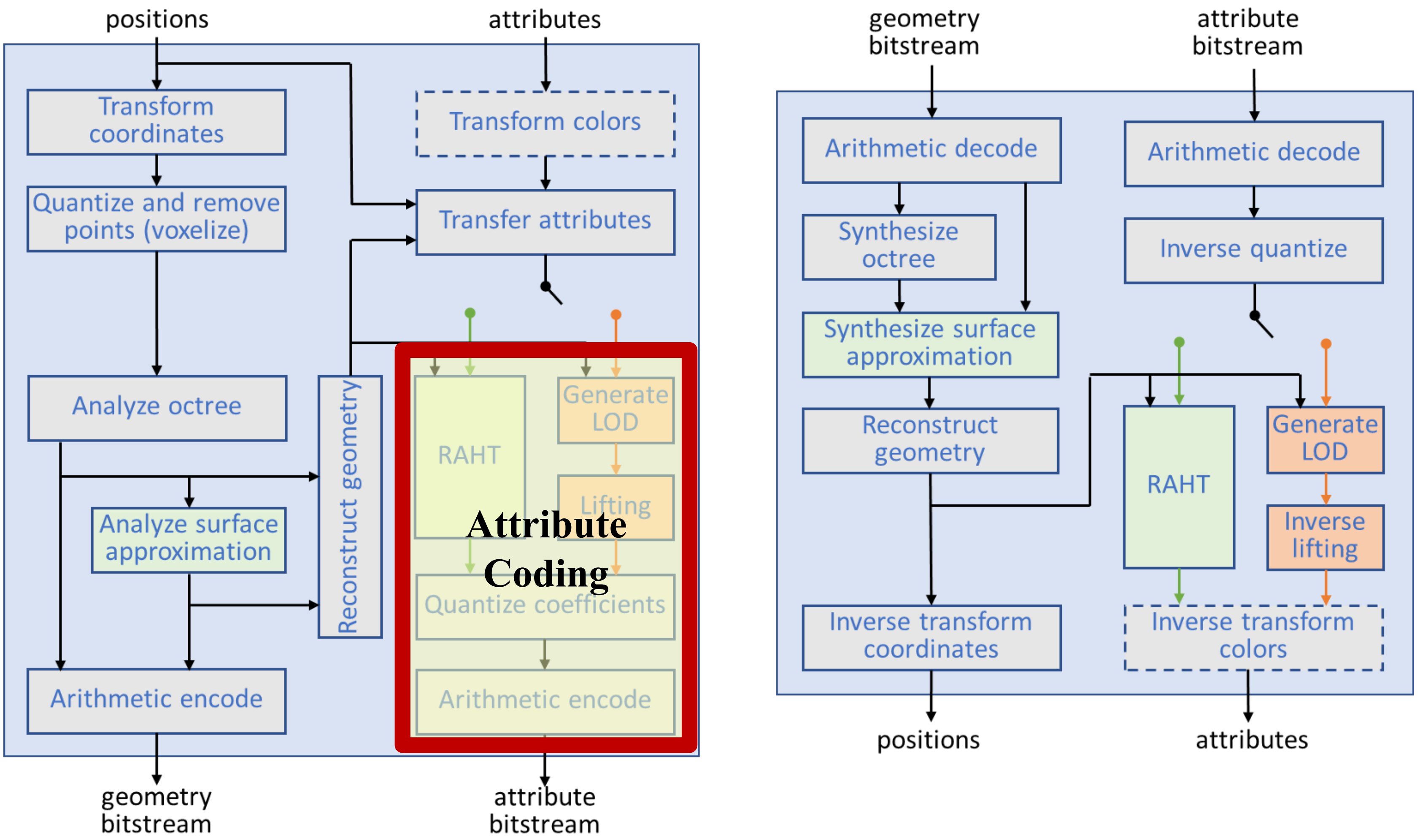


Re-coloring

- **Re-coloring:** assigns colors from the original point cloud to the decoded points
  - Not needed if lossless geometry coding is performed
- Each decoded point is assigned with the color of the closest input point in terms of Euclidean distance
- Significant impact on the color coding efficiency!

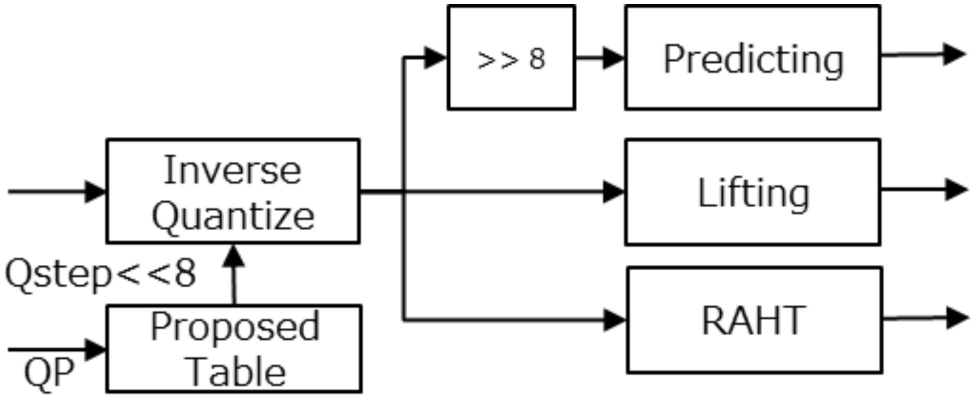
S. Kuma, O. Nakagami, "PCC CE1.3 Recolor method", ISO/IEC M42538, San Diego, USA.

G-PCC Architecture



G-PCC Attribute Coding Modes

- G-PCC defines 3 attribute (color) coding methods:
  - **RAHT transform:** Region Adaptive Hierarchical Transform (RAHT) coding
  - **Predicting transform:** interpolation-based hierarchical nearest-neighbour prediction
  - **Lifting transform:** interpolation-based hierarchical nearest-neighbour prediction with an update/lifting step
- RAHT and Lifting are typically used for Category 1 data, while Predicting is typically used for Category 3 data
  - User has the option to choose which of the 3 attribute codecs should be used



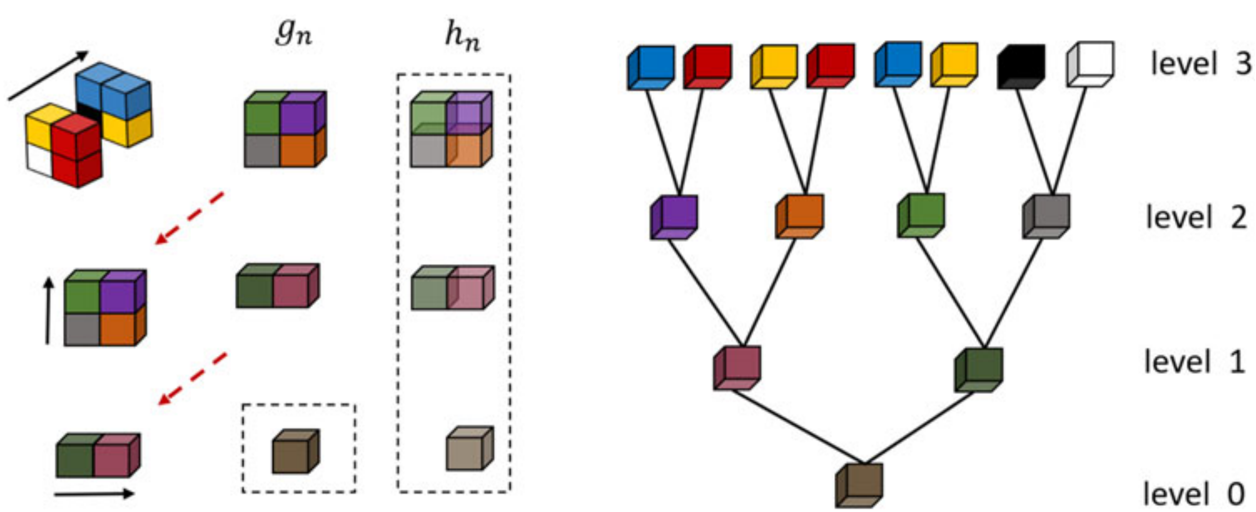
from 3DGroup, "G-PCC codec description v3" N18486, April 2018.

Color Coding: Region Adaptive Hierarchical Transform (RAHT)

- RAHT transform is is essentially an Haar transform with the coefficients appropriately weighted
- RAHT is performed with respect to a hierarchy defined by the so-called Morton codes (i.e. indices) of the voxels
  - 3D transform where the three dimensions are processed one after the other
  - Recursively applied from the target depth up to the octree root, to blocks (of voxels) grouped two by two at each level
  - Hierarchical since the transform in one dimension is applied to the coefficients resulting from the transform in the previous dimension
- **Quantization:** transform coefficients are quantized by a uniform scalar quantizer
- **Arithmetic coding:** quantized transformed coefficients are coded using a binary arithmetic encoder

RAHT Implementation

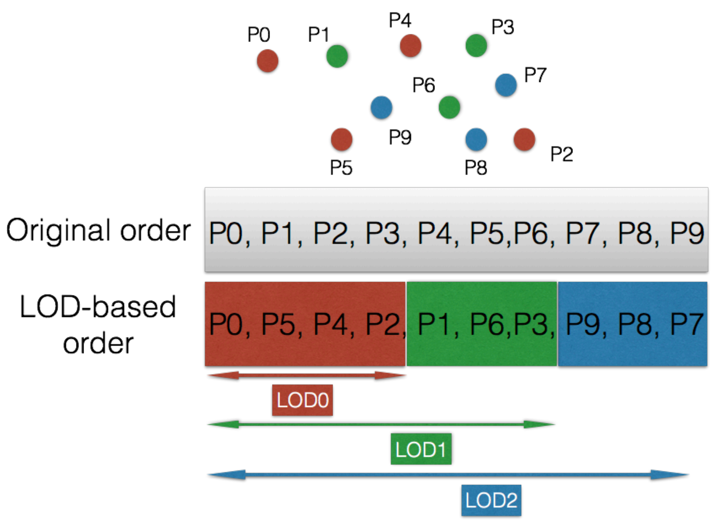
- RAHT produces low-pass and high-pass coefficients at each decomposition
  - Fast fixed-point implementation is available
- Transform matrix is adaptive based on the number of leaf voxels that each coefficient represents
  - Takes into consideration empty voxels
- Transform domain prediction of RAHT coefficients from spatially up-sampling the DC values of neighboring parent nodes



From D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki and A. Tabatabai, "An Overview of Ongoing Point Cloud Compression Standardization Activities: Video-based (V-PCC) and Geometry-based (G-PCC)", APSPIPA, March 2020.

Color Coding: Prediction Transform

- **Prediction scheme:** already encoded colors are considered for prediction
  - Points encoding order is very important for coding efficiency
- Defines two main steps:
  - **Level of Details (LoD)** generation process defines an efficient encoding order
    - Predictors can come from the previous or the same LoD
  - **Linear interpolation** process based on the distances to the already encoded k-nearest neighbors of the point
- Based on the LoD structure, the predicting transform is applied with a split and prediction step applied sucessively



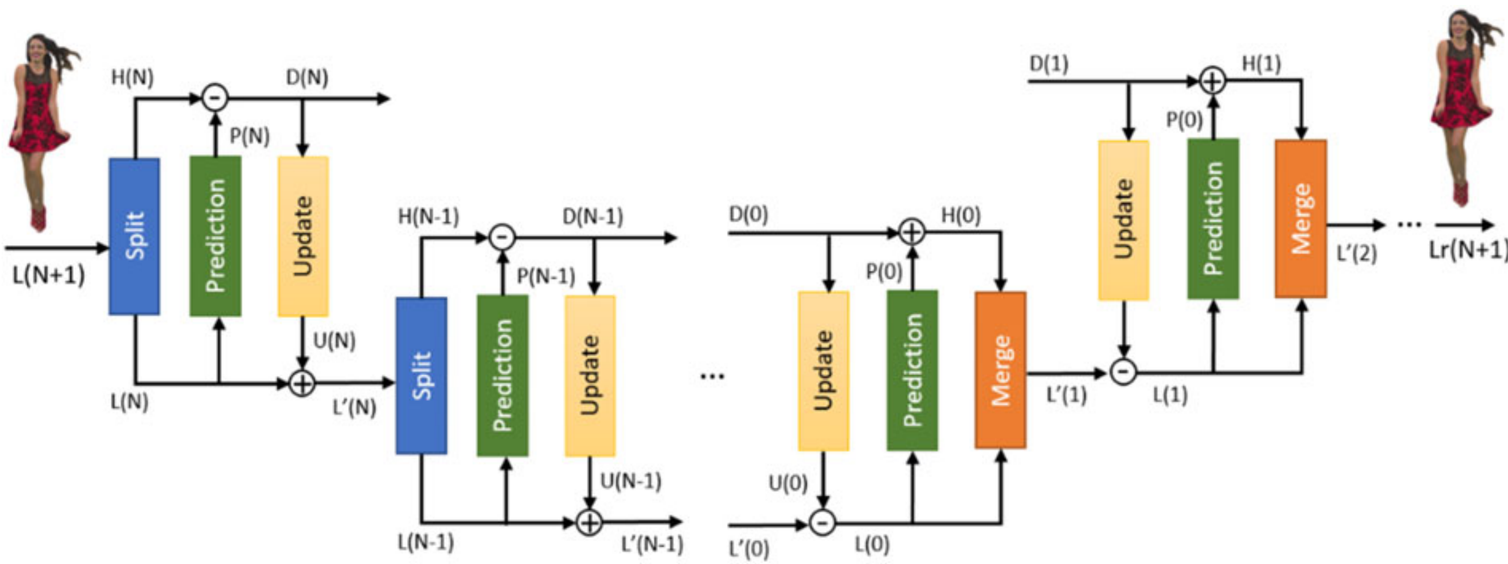
LOD re-organizes the points into a set of refinement levels based on the Euclidean distances between points

from 3DGroup, "G-PCC codec description v3" N18486, April 2018.

Color Coding: Lifting Transform

- Improvement to the Predicting Transform with two additional tools:
  - Update operator
  - Adaptive quantization

- Leverages the prediction residuals to update the color/attribute values at each stage of the prediction and guide the quantization process

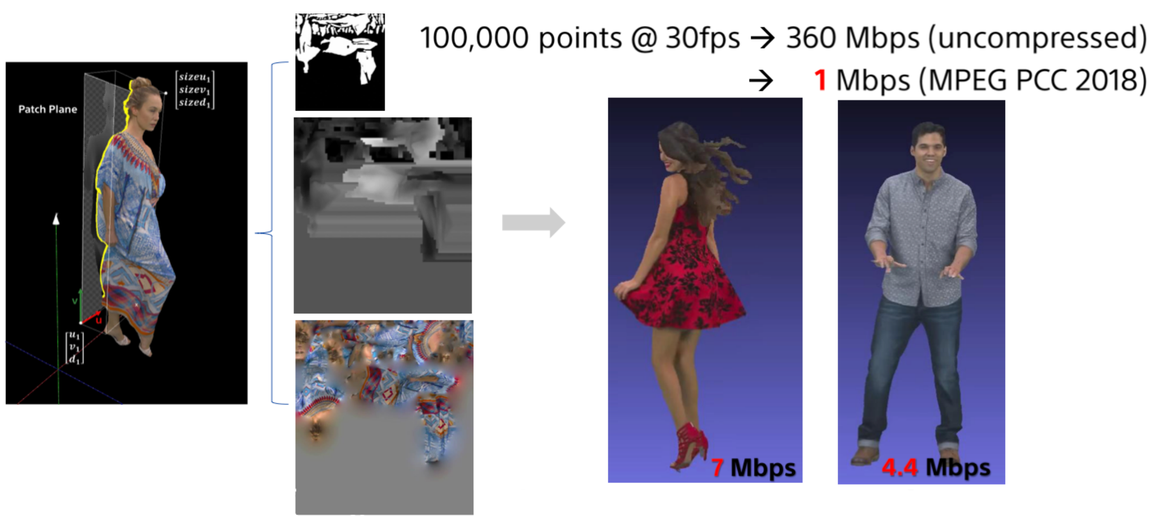


From D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki and A. Tabatabai, "An Overview of Ongoing Point Cloud Compression Standardization Activities: Video-based (V-PCC) and Geometry-based (G-PCC)", APSPIPA, March 2020.

Point Cloud Coding:
MPEG V-PCC

V-PCC (ISO/IEC 23090-5) Overview

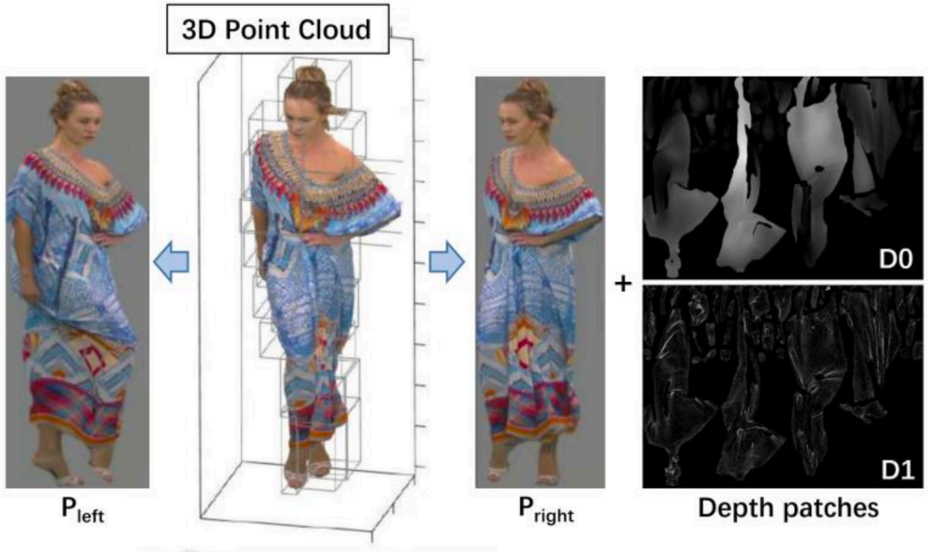
- **MPEG Video-based Point Cloud Compression (V-PCC):**
  - Reached Draft International Standard (DIS) in March 2020
- Targets **dynamic PC coding:** suitable for **dense point cloud animations** that generate video sequences with well defined surfaces
  - Combination of 2D and 3D compression technologies
  - Leveraging on decades of 2D video technology development
  - Supports lossy and lossless coding



from Danillo Graziosi

V-PCC Principles and Foundations

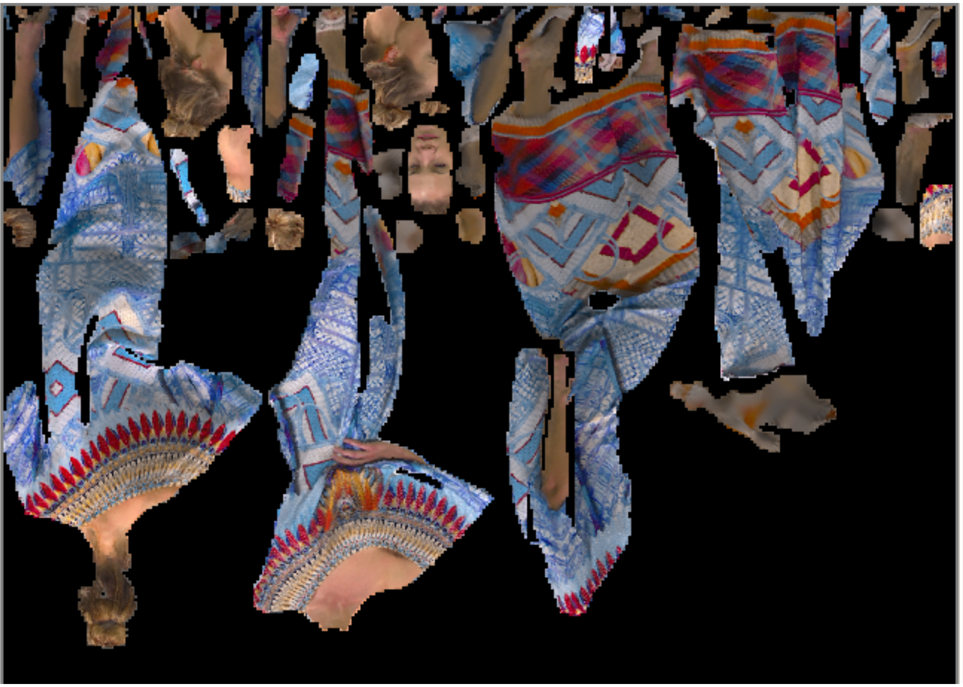
- **MPEG V-PCC 3D to 2D mapping process:** generates planar patches using orthographic projection to different sides of a bounding box
- Neighboring points in the XY plane are grouped as a patch and each patch may be represented by two different images:
  - Geometry image: grayscale image which contains the Z coordinate information
  - Texture image: three component image which contains the color information



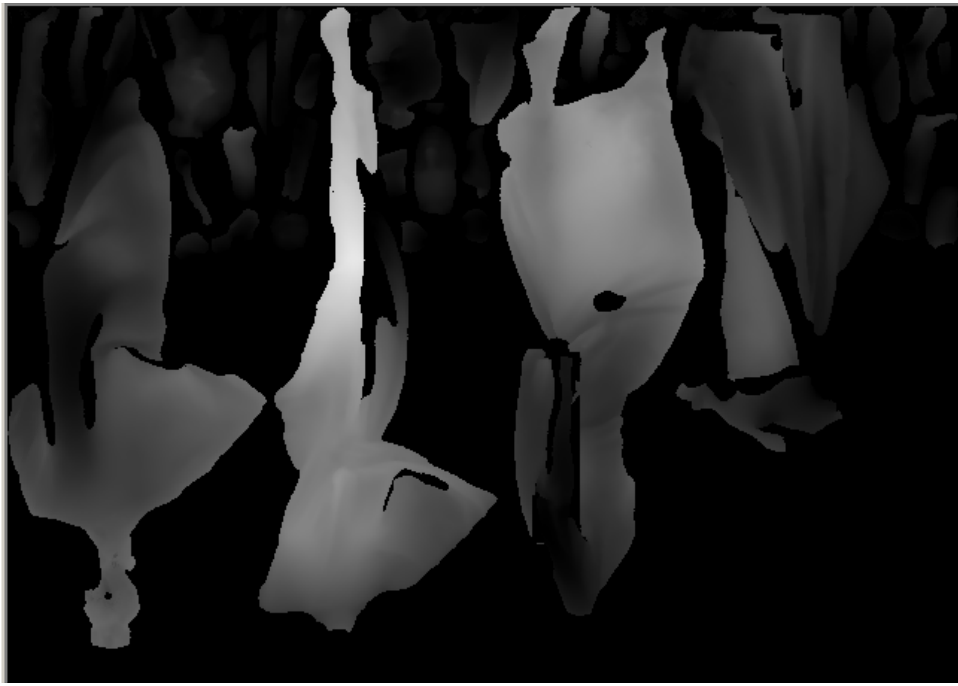
G. Lafruit, A. Schenkel, C. Tulvan, M. Preda, Lu Yu, "MPEG-I coding performance in immersive VR/AR applications", IBC 2018 Conference, Sep 2018, Amsterdam, Netherlands.

V-PCC Point Cloud Data Representation

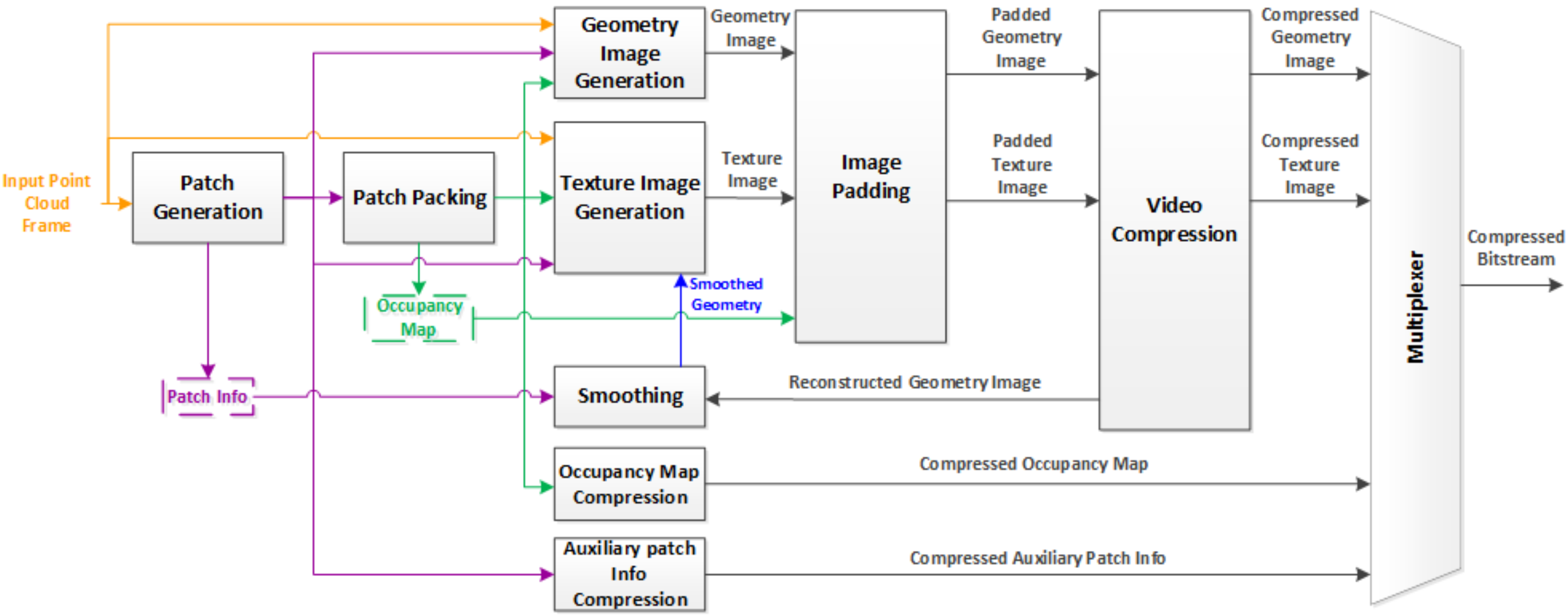
Texture



Geometry

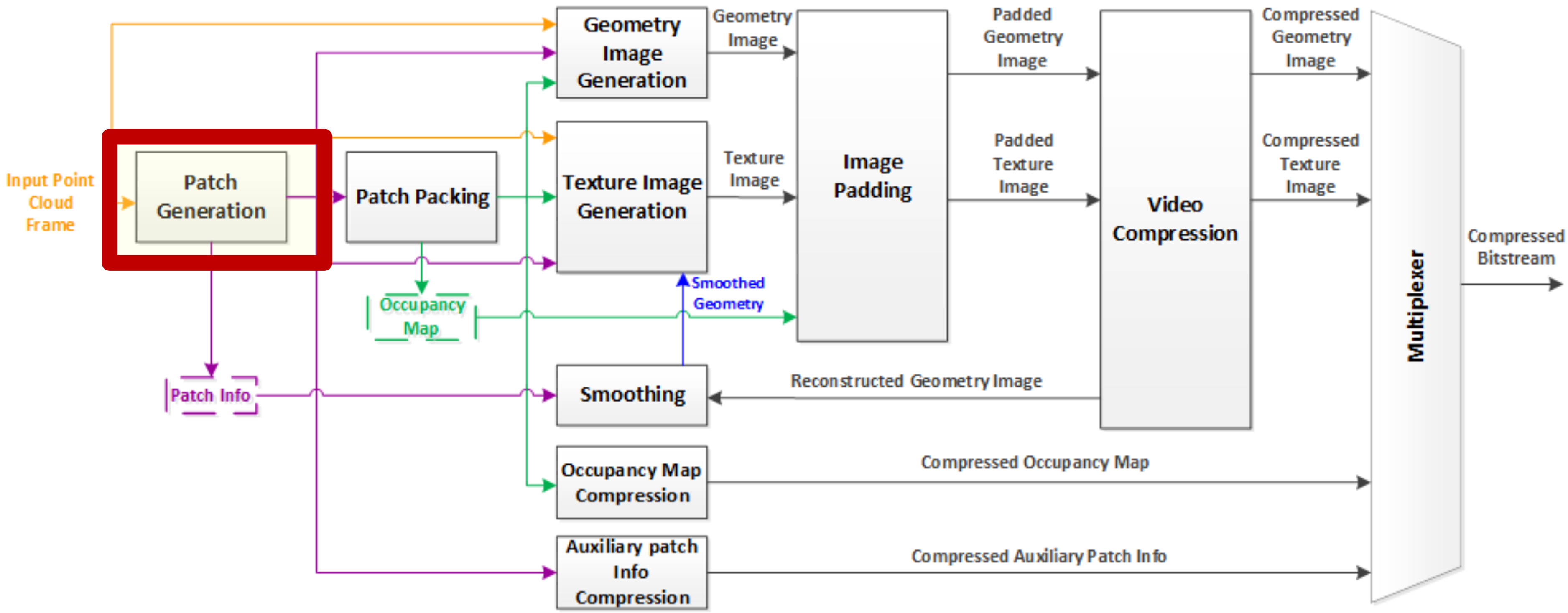


V-PCC Architecture

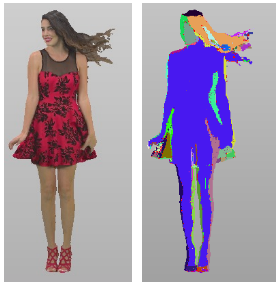


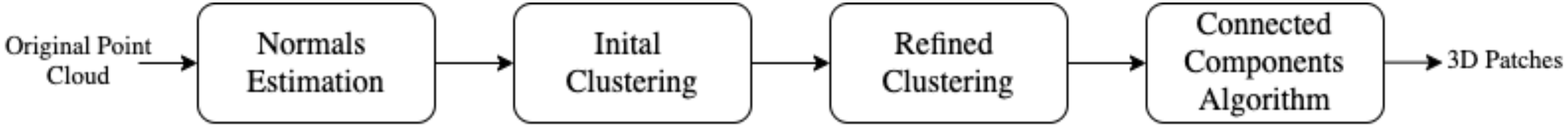
from V. Zakharchenko, "Algorithm description of mpeg-pcc-tmc2," N17767, Ljubljana, Slovenia, Jul. 2018

V-PCC Architecture



V-PCC Patch Generation





- **Normals Estimation:** normals are estimated for each point, e.g. by computing PCA over the 16 nearest neighbors and the point
- **Initial Clustering:** normals are clustered into six clusters according to the maximization of the dot product to one of six pre-defined planes (X,-X,Y,-Y,Z,-Z)
- **Refined Clustering:** clustering is refined with some spatial criteria, e.g. a point may be assigned to another cluster based on the neighboring points clusters
- **Connected Components Algorithm:** clusters are segmented into 3D patches (or sub-clusters) of points that are spatially connected with a connected components algorithm

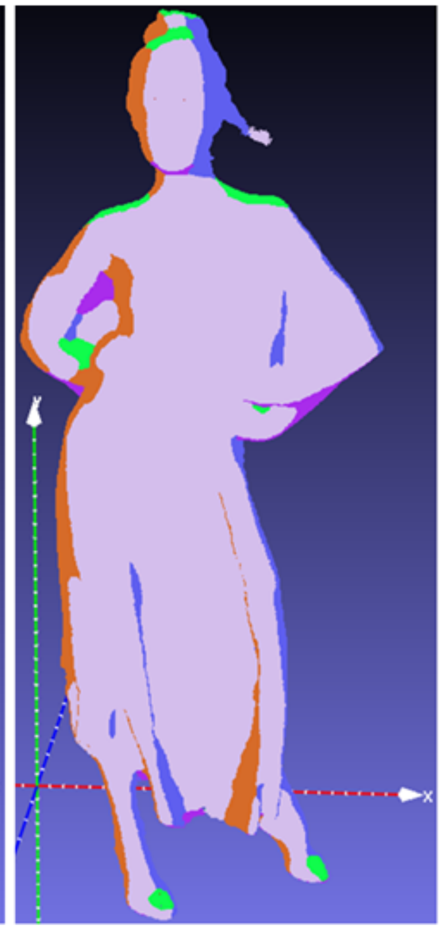
TÉCNICO LISBOA

V-PCC Patch Generation Examples

Original PC

Initial Clustering

After Refinement



Final Clustering

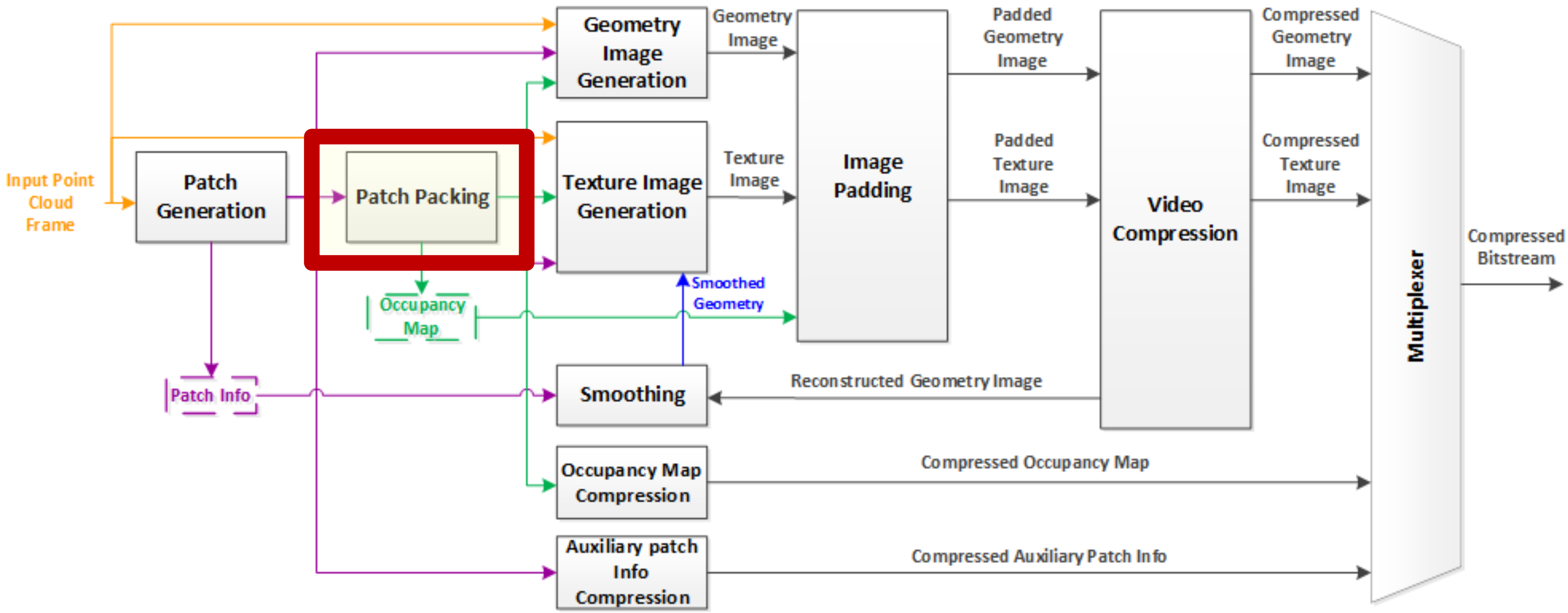
Patch Generation Iterative Algorithm

- V-PCC patch generation follows an iterative algorithm where clustering is refined and connected components are established several times
  - Points within a 3D patch must have low depth variance (easy to code)
  - Points may be removed if their depth value is out of a specified range
  - 3D patches cannot be too small
  - Removed points can be re-clustered again so that new 3D patches can represent them
- At the end, the total number of clustered points could be lower than the original points

V-PCC Patch Projection

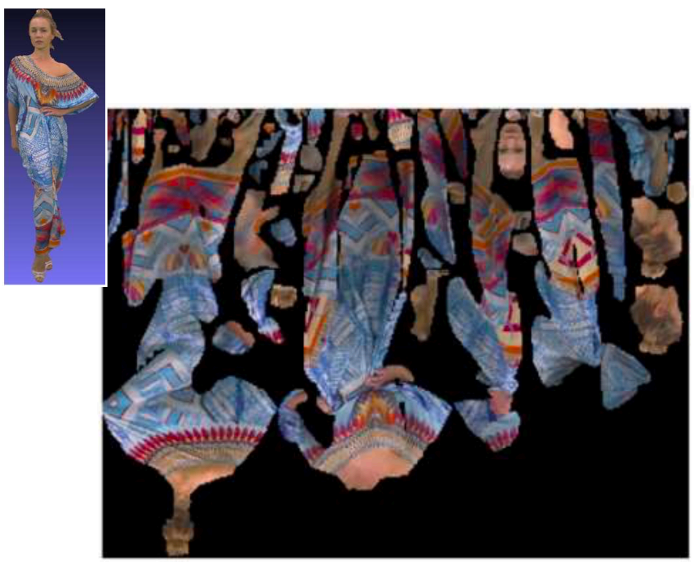
- Orthographic projection into six faces of the bounding box (projection planes) is applied to each 3D patch, to avoid any resampling issues
  - Each 3D patch has its own 3D local coordinate system defined by δ (depth), s (tangent) and r (bi-tangent) planes, where the depth axis is aligned with the normal of the projection plane
- Since the use of only six projection planes can be limited, 12 new modes corresponding to cameras at 45º were added
  - Point cloud is rotated for these new modes (only integer operations are made)
- A map with several layers can be used to store projected points, most often two layers (D0 and D1) to accommodate possible occluded points
  - Difference between the same pixel on D0 and D1 layer can only be up to some surface thickness threshold (usually 4)
  - When multiple points fall into the same location:
    - The one with lowest depth falls on D0
    - The one with second lowest depth falls on D1
    - The other points are simply not projected

V-PCC Architecture



V-PCC Patch Packing

- **Packing:** process to place the projected 3D patches into a 2D image
- **Packing strategy:** insert patches in a W×H image minimizing the unused space
  - Every T×T map block (e.g. 16×16) is associated with one single patch
- **Occupancy map:** created during packing with information about each T×T block
  - Binary map to indicate if the block belongs or not to a point cloud patch

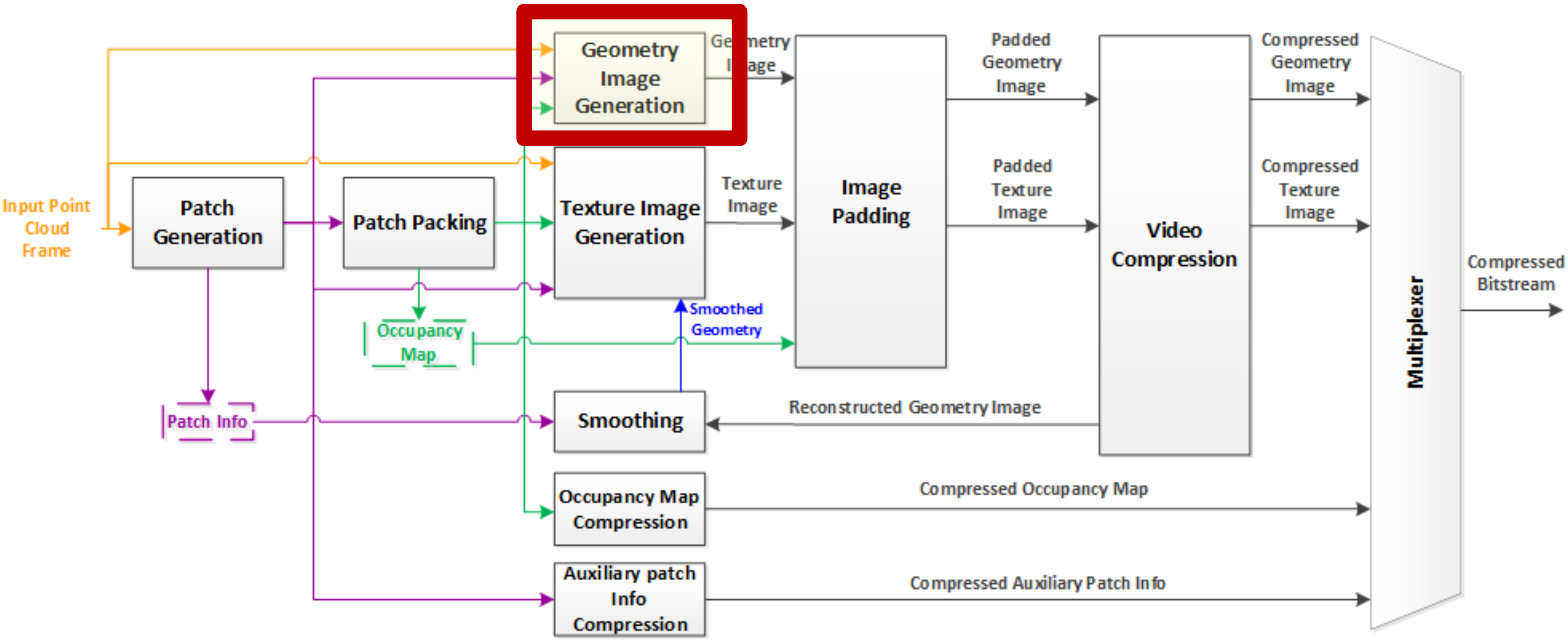


V-PCC Patch Packing Strategy

- Packing strategy is a **non-normative tool**!
- Patches are **sorted** by their size and inserted individually by searching an empty space in the image that can fit the patch (first empty space is selected)
  - Raster scan order search
- Patches can be **rotated** and **mirrored** to make the packing more compact
- After insertion, the blocks used by the patch are marked as non-empty
  - Guarantees an overlapping-free insertion
  - Includes a mechanism to increase map height if more space is needed
- **Consistent packing over time:** for each current frame patch, the best matching patch in the previous frame is searched
  - Sorting criteria: ratio between the intersection area over the union area of the patch bounding boxes in both reference and current frames

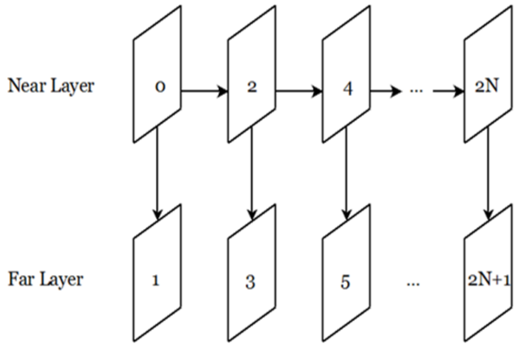
TÉCNICO LISBOA

V-PCC Architecture



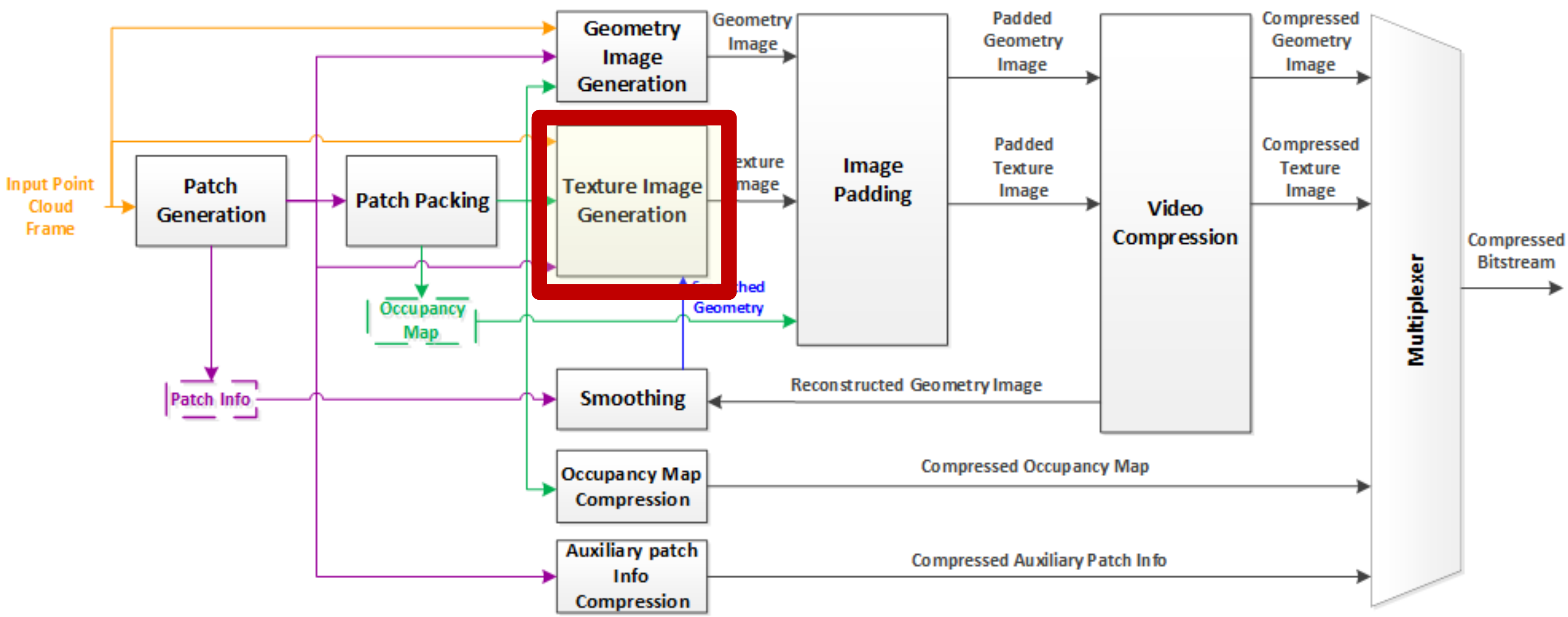
Geometry Image Generation

- Geometry images (also called depth maps) are **luminance only** and represent the distance between the 3D point and the corresponding projection plane
  - X and Y 3D coordinates correspond to the position of the pixels within the frame
  - Z 3D coordinate (depth) corresponds to the pixel intensity value
- **Several 3D points may correspond to the same mapped pixel** which can be handled with the two D0 (near-layer) and D1 (far-layer) map layers
  - For example, the outside and inside of an object
- Several options are possible to code D0 and D1 layers:
  - D1 is coded differentially with respect to D0
  - Temporally interleaving of D1 and D0
  - D0 and D1 can be sub-sampled and spatially interleaved
  - D1 can be dropped and interpolated at the decoder



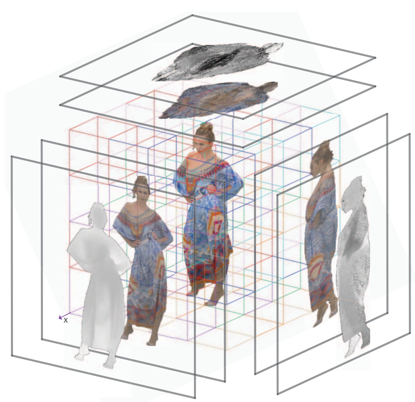
TÉCNICO LISBOA

V-PCC Architecture



Texture Image Generation

- Since geometry data may be lossy, attribute image generation requires the decoded geometry to compute new colors (to be transmitted) associated to the decoded points
- **Re-colouring process** transfers the color information from the original point cloud to the decoded geometry point cloud. The following data is used:
  - Color value of the nearest point in the original point cloud
  - Color values of the nearest neighbor points in the decoded point cloud
- The pixels in the attribute images take the color values of the mapped points



From L. He, W. Zhu and Y. Xu, "Best-Effort Projection Based Attribute Compression for 3D Point Cloud," 23rd Asia-Pacific Conference on Communications (APCC), Perth, WA, Australia, Dec. 2017.

V-PCC Architecture

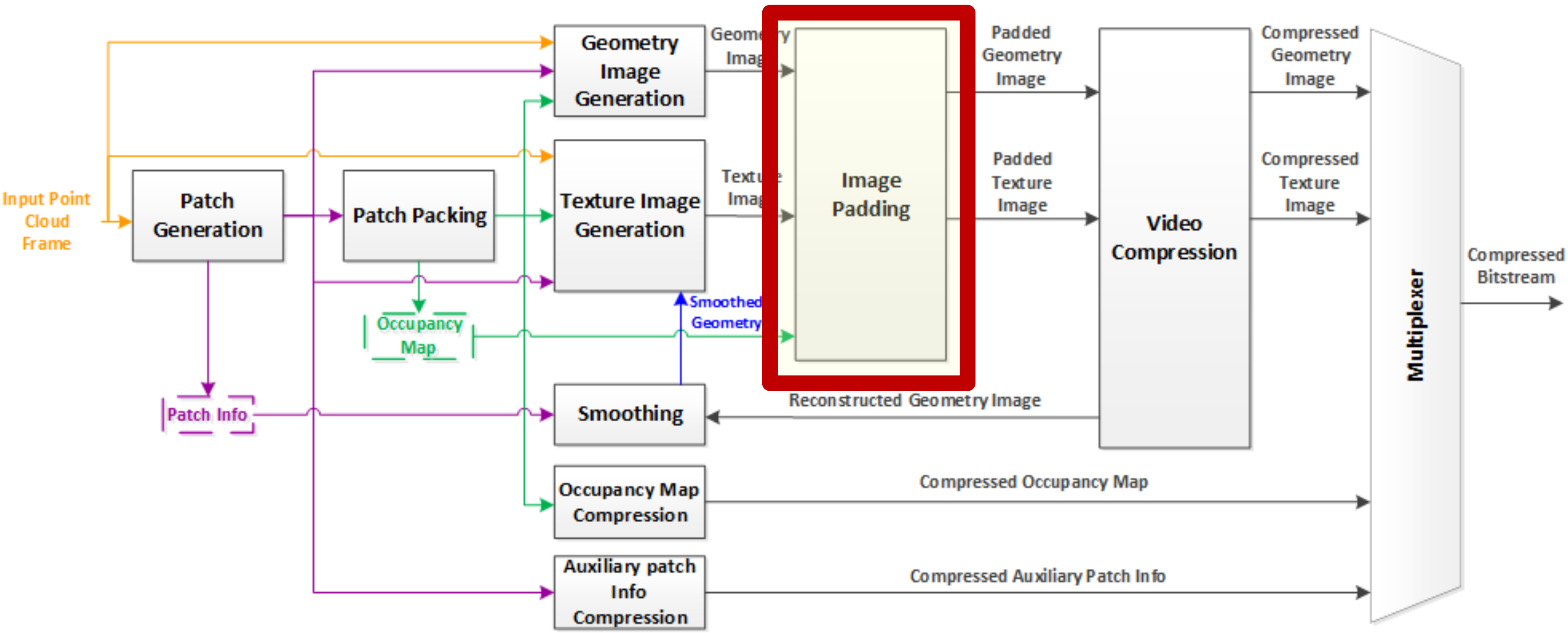


Image Padding

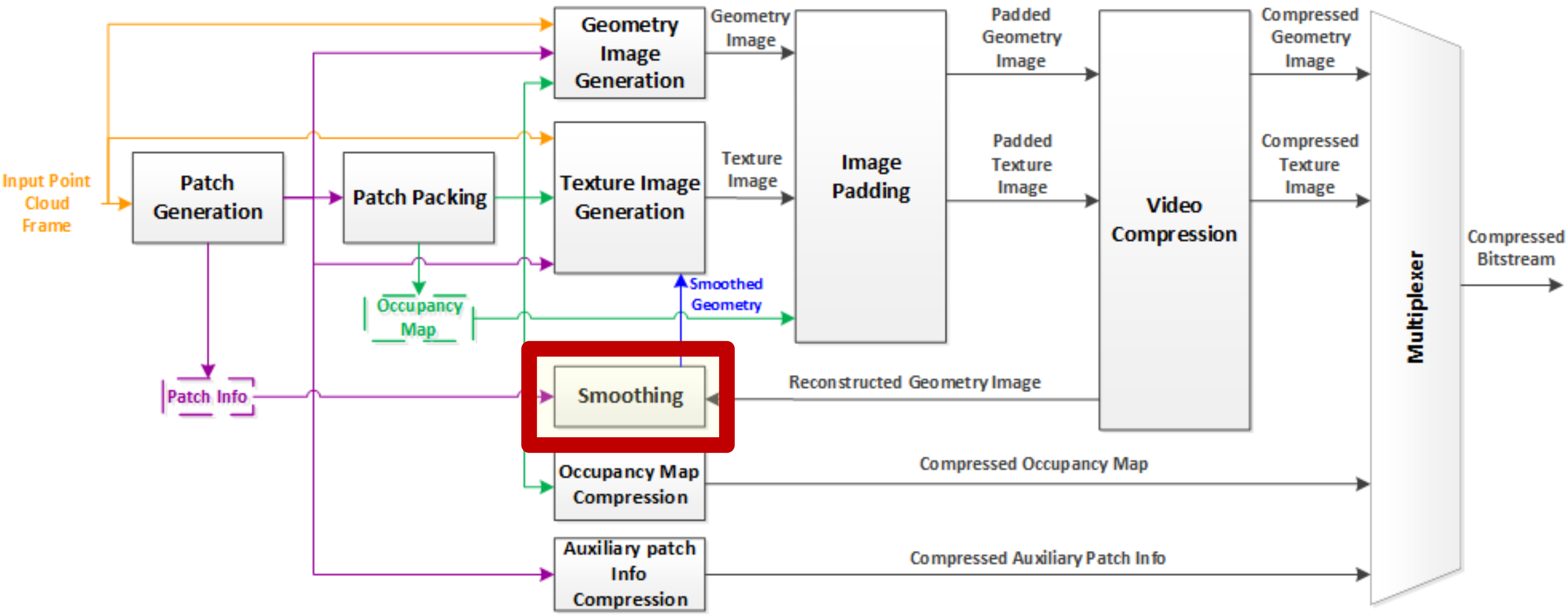
- To **enable an efficient video compression and minimize block artifacts** images must not have high frequencies
  - Empty space between patches must be filled/padded appropriately
- For geometry images, an **adaptive padding strategy** is performed at T×T block level:
  - Block with no empty pixels: nothing is done
  - Empty block: pixels in the block are filled with the values of the last block row/column
  - Partially filled block: empty pixels are iteratively filled with average values of neighboring pixels
  - Group dilation is performed to maintain consistency between the D0 and D1 layer maps
- For texture images, an optimization procedure finds the values of empty pixels such that the obtained padded image is as smooth as possible

Image Padding Example

Before Padding

After Padding

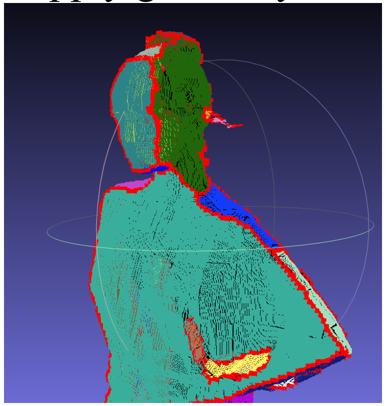
V-PCC Architecture



Attribute and Geometry Smoothing

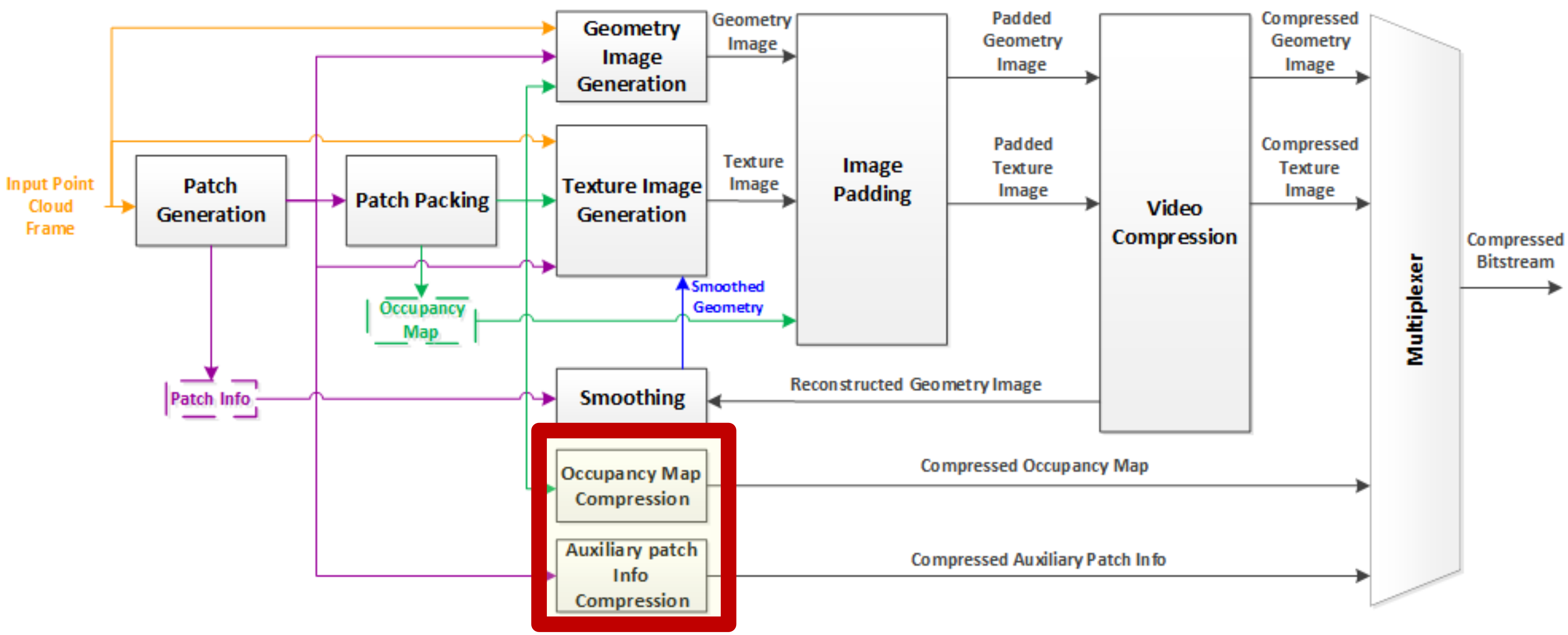
- Several types of **coding artifacts** may appear due to**:**
  - Lossy compression of geometry and texture images
  - Occupancy map subsampling
  - Segmentation into patches
- **Geometry smoothing** avoids discontinuities that may appear at the patch boundaries
  - Boundary points are moved to the centroid of their nearest neighbors (trilinear filter on a 2x2x2 grid neighborhood).
- **Attribute smoothing** can also be applied between patch boundaries to reduce the appearance of seams
- Smoothing is signaled to the decoder

Points to apply geometry smoothing



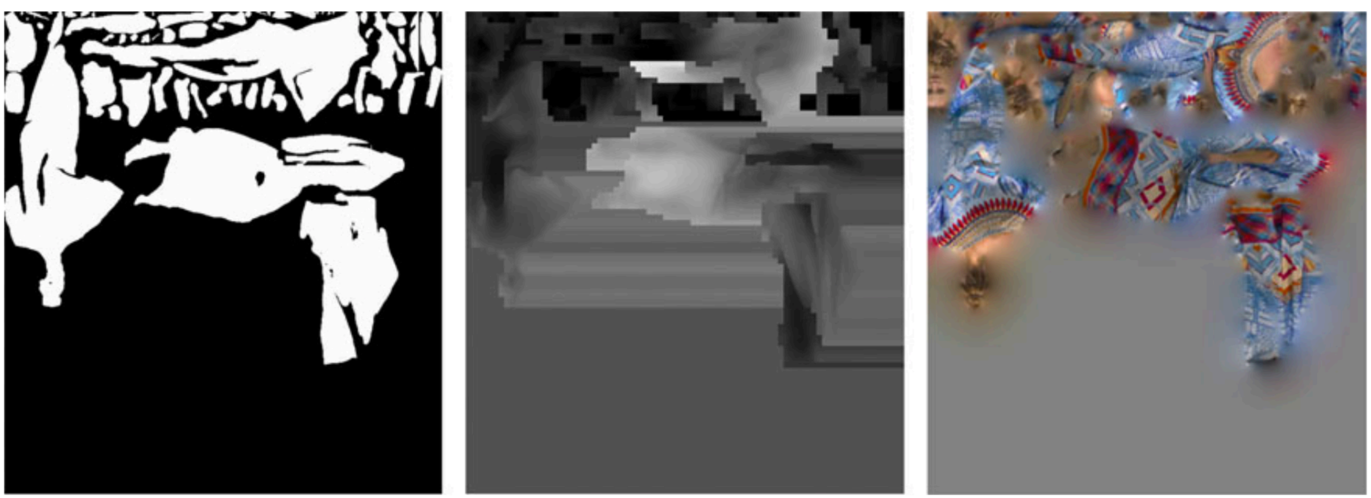
O. Nakagami, "PCC TMC2 low complexity geometry smoothing," ISO/IEC m43501, July 2018, Ljubljana, SI

V-PCC Architecture



Occupancy Map Compression

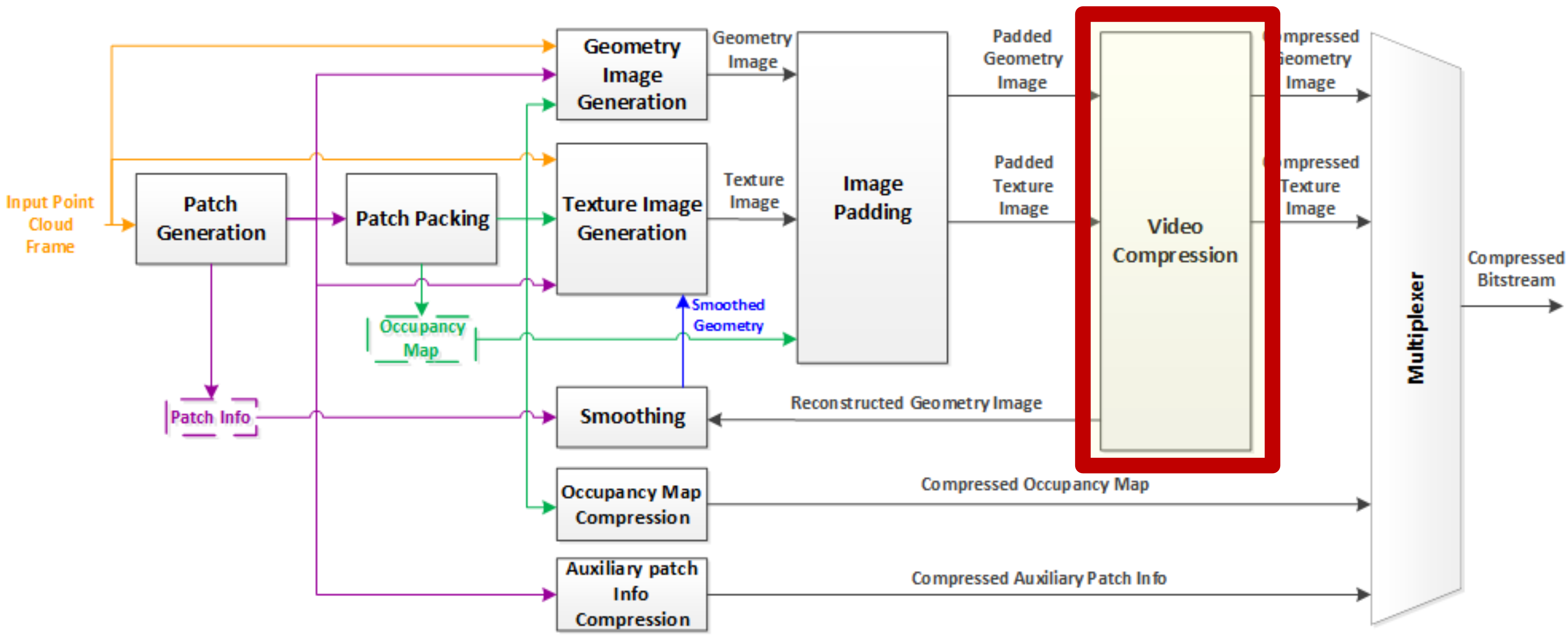
- **Occupancy map** is a binary image coded using a lossless video encoder with a precision of $B_0 \times B_0$ blocks
  - Each $B_0 \times B_0$ block is represented with 1 if it contains at least a non-padded pixel and 0 otherwise
  - B_0 is a user-defined parameter which can control between precision and rate cost
- The occupancy map can also be coded as a down-sampled video frame, using a video codec with lossy support



Auxiliary Data Information Compression

- For the decoder to perform the 2D to 3D mapping, for every patch, **auxiliary data** is encoded and transmitted:
  - Index of the projection plane associated to its normal direction
  - 2D bounding box
  - 3D location
- The **index of the unique patch** (in an ordered list) associated to each T×T block is also encoded
- For the case of a **temporally consistent packing,** this information can be differentially encoded using the values of the matching patches from a previous frame
- All information is signaled as HEVC NALs and follows a similar syntax

V-PCC Architecture



2D Video Compression

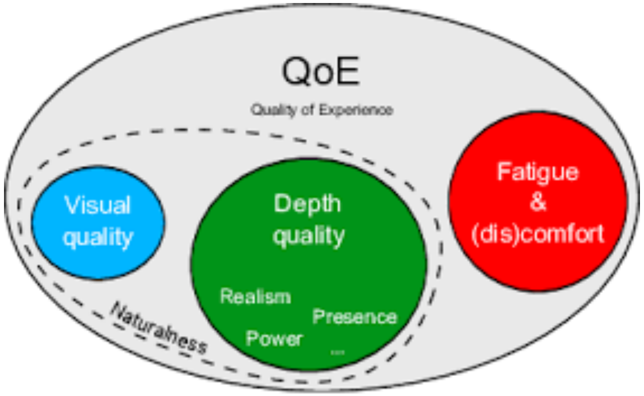
- The successive generated texture, geometry and occupancy maps can be coded as video frames using any **video coding standard**
  - HEVC is now the most popular choice
  - 10-bit profiles are preferred since they can improve accuracy and coding efficiency
  - Motion estimation process can be improved with patch location placement
- At the decoder side, **four different coded streams** are demultiplexed and decoded:
  - Patches information, the occupancy map and the geometry images are decoded first
  - Reconstructed points with only geometry is obtained
  - Texture information obtained at the same position from the decoded texture image
- The **number of decoded points** may be much larger than the number of original points
  - Controlled by the occupancy map coding precision B_0
  - More dense point clouds are obtained and thus better perceptual quality

5 Point Cloud Quality Assessment

Quality of Experience

- Quality of Experience (QoE) “is the degree of **delight** or **annoyance** of the user of an **application** or **service**. It results from the fulfillment of his or her **expectations** with respect to the utility and/or enjoyment of the application or service in the light of the **user’s** personality and current state”

- QoE assessment is fundamental to evaluate the performance of the several point cloud processing steps:
  - Such as denoising, coding and rendering



from “Qualinet white paper on definitions of quality of experience”, European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, 2012.

User Quality: Mostly Signal Fidelity

- Subjective Evaluation

- Objective Evaluation

$$\mathrm{PSNR} = 10 \times \lg\left(\frac{255^2}{\mathrm{MSE}}\right)$$

$$\mathrm{MSE} = \frac{1}{M \times N}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[I(i,j) - I'(i,j)\right]^2$$

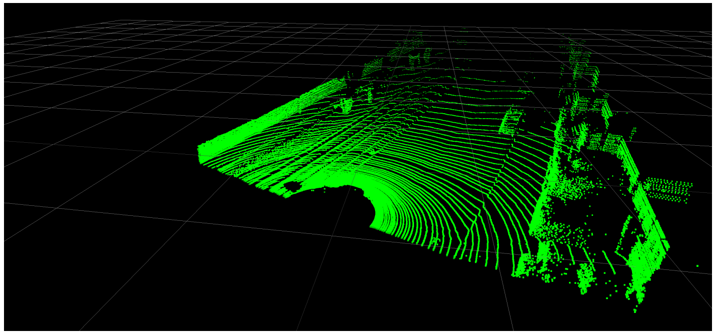
What is Mean Opinion Score (MOS)?

- The **likely level of satisfaction** of a specific service/product dimension, e.g. visual quality, as appreciated by an average user (from a provider point of view)

- Should be performed such that it **generates reliable and reproducible results**
  - Subjective evaluation methodology
  - More complex and difficult that it *a priori* seems
  - Much used for video and audio subjective qualities

Common Test Conditions (MPEG)

- Well-defined test environment to conduct experiments and thus, compare the outcome of experiments
- Three test categories with several test conditions
  - **Category 1**: Static Objects and Scenes
  - **Category 2:** Dynamic Objects
  - **Category 3:** Dynamic Acquisition
- **Lossy** and **lossless geometry** coding with attributes (or not) lossy or lossless coded
- Defines parameterization of G-PCC and V-PCC codecs
- Defines rate/quality metrics: **PSNR D1 and PSNR D2**



TÉCNICO LISBOA

5.1 Subjective Quality Assessment

Subjective Quality Assessment

- Subjective tests aim at producing **User Opinion Scores** as a delicate mixture of ingredients and choices:
  - Test & lab environment
  - Test material
  - Test methodology
  - Test subjects
  - Data analysis

PC Subjective Quality Assessment

How to measure point cloud quality in a subjective way?

- Different alternatives for subjective test methodologies (and grading scale):
  - Double-Stimulus Impairment Scale (DSIS)
  - Absolute Category Rating (ACR)
  - Pairwise comparison (PC)
- Different alternatives for display:
  - Augmented (or virtual) reality headset
  - Standard 2D displays
  - Stereoscopic and auto-stereoscopic displays
- Different alternatives for interaction:
  - With content interaction
  - With no-content interaction
- Different alternatives for rendering:
  - Point-based rendering (with color or not)
  - Mesh-based rendering

Octree Coding Artifact Characterization

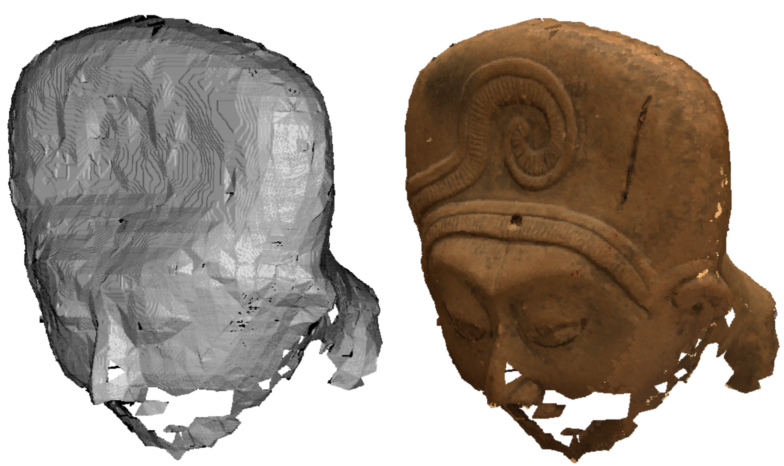
- As the target bitrate decreases (lower octree depth), the number of decoded points also decreases
  - All points inside a voxel are represented by just one point at the voxel center
- Some types of **artifacts**:
  - Lack of detail
  - Pixelation (sub-sampling)
  - Annoying false boundaries

Point-based rendering

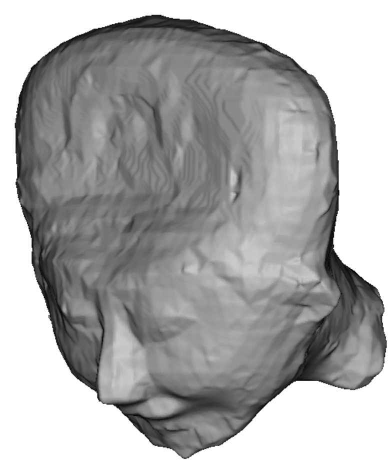
Mesh-based rendering

G-PCC Trisoup Coding Artifact Characterization

- Number of points no longer reduced
- Trisoup process create false edges at the boundaries of the blocks or triangles
- Some types of **artifacts**:
  - Artificial holes with polygonal shapes
  - Triangles are now visible
  - Lost detail in some regions



Point-based rendering



Mesh-based rendering

V-PCC Coding Artifact Characterization

- V-PCC is rather efficient to hide annoying artifacts and large deformations on the geometry typically do not occur
- Traditional (HEVC) block-based tools are responsible for the introduction of the artifacts
- Some types of **artifacts**:
  - Artificial holes with polygonal shapes
  - Missing points due to the way the projection is made

Point-based rendering

Mesh-based rendering

G-PCC/V-PCC Coding Efficiency Study I

From E. Alexiou et al., “A comprehensive study of the rate-distortion performance in MPEG point cloud compression”, APSIPA Transactions on Signal and Information Processing, Vol. 8, 2019.

Content	Repository	Pre-processing	Voxelization	Voxel depth	Input points	Output points
Objects						
amphoriskos	Sketchfab	✓	✓	10-bit	147.420	814.474
biplane	JPEG	✗	✓	10-bit	106.199.111	1.181.016
head	MPEG	✗	✓	9-bit	14.025.710	938.112
romanoillamp	JPEG	✓	✓	10-bit	1.286.052	636.127
Human figures						
longdress	MPEG	✗	✗	10-bit	857.966	857.966
loot	MPEG	✗	✗	10-bit	805.285	805.285
redandblack	MPEG	✗	✗	10-bit	757.691	757.691
soldier	MPEG	✗	✗	10-bit	1.089.091	1.089.091
the20smaria	MPEG	✗	✓	10-bit	10.383.094	1.553.937

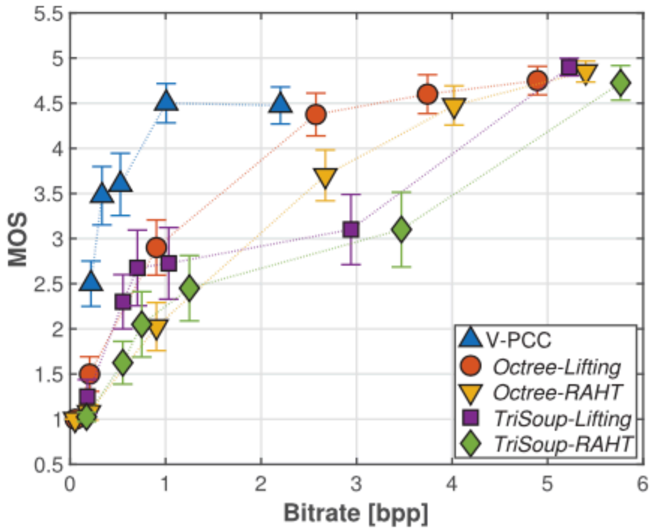
(a) *amphoriskos* (b) *biplane* (c) *head* (d) *romanoillamp*

(e) *longdress* (f) *loot* (g) *redandblack* (h) *soldier* (i) *the20smaria*

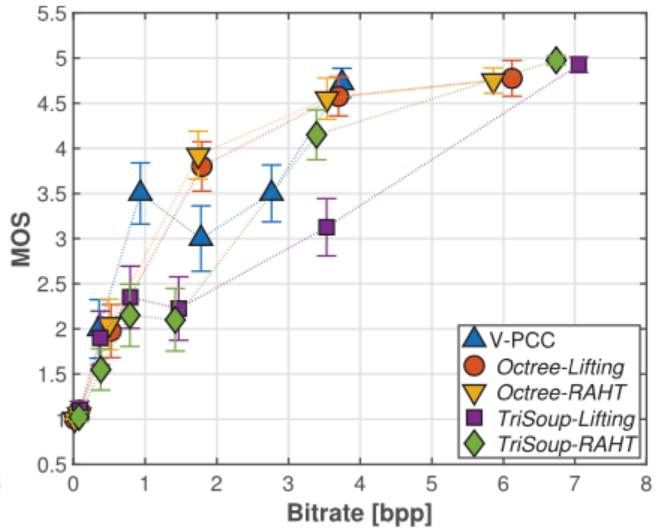
G-PCC/V-PCC Coding Efficiency Study II

- Coding configurations under evaluation:
  - V-PCC geometry and color coding
  - G-PCC Octree geometry coding: lifting and RAHT color coding
  - G-PCC Trisoup geometry coding: lifting and RAHT color coding
- Six degradation levels for G-PCC encoder and 5 degradation levels of V-PCC encoder
  - Parameterizations follow the MPEG test conditions
- Subjective evaluation experiments in two laboratories MMSPG at EPFL in Lausanne, Switzerland and LISA at UNB in Brasilia, Brazil
  - Controlled environment following ITU-R BT500.13
  - Apple Cinema Display of 27-inches and 2560 × 1440 resolution
  - Point-based rendering with adaptive point size
  - Double-Stimulus Impairment Scale (DSIS) with 5-grading scale
  - Total of 40 subjects with 2 sessions (10 min break)

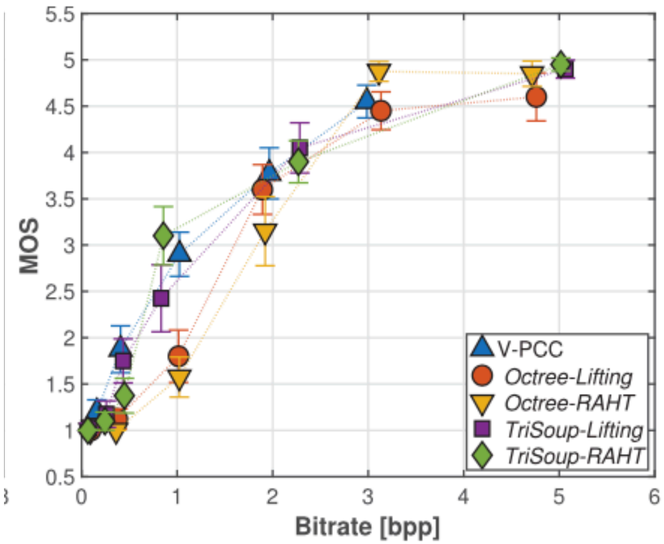
Rate-MOS Quality Evaluation I



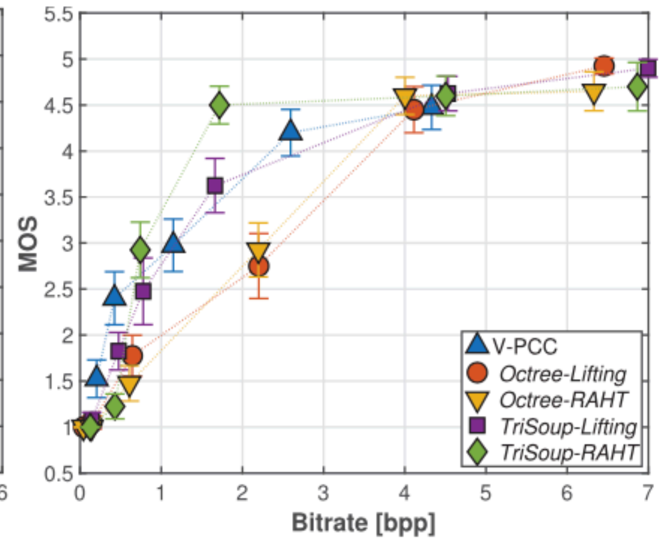
(a) *amphoriskos*



(b) *biplane*

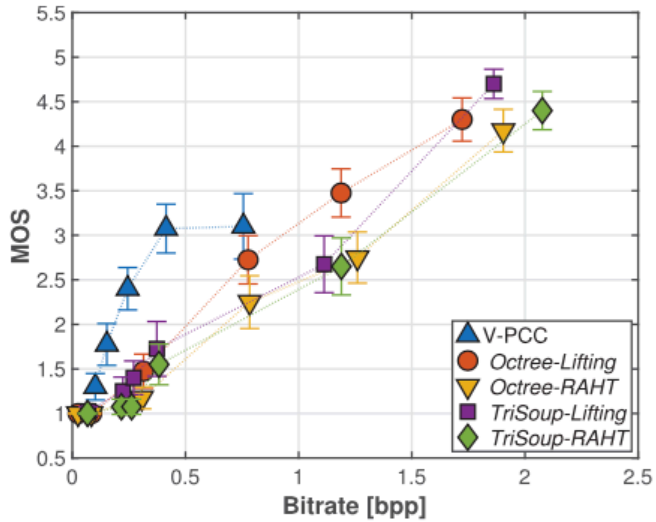


(c) *head*

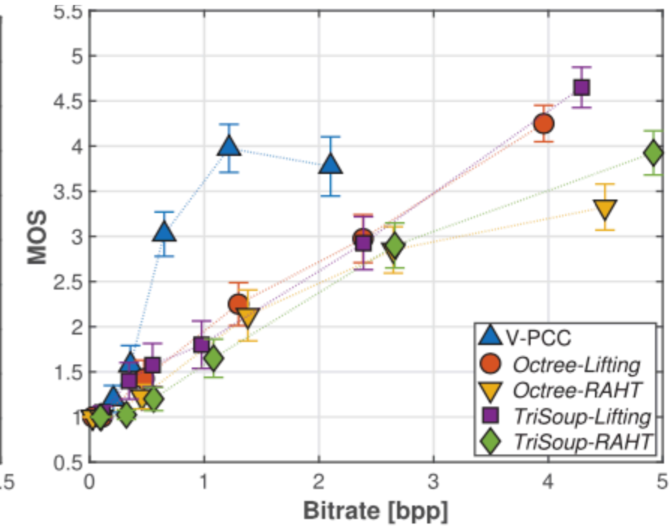


(d) *romanoillamp*

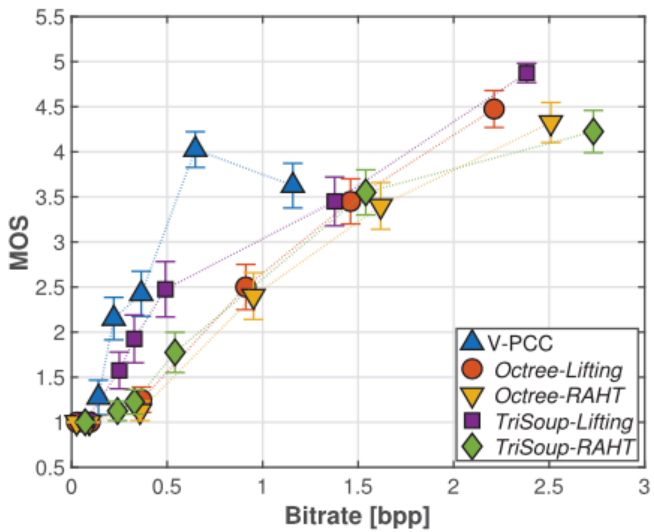
Rate-MOS Quality Evaluation II



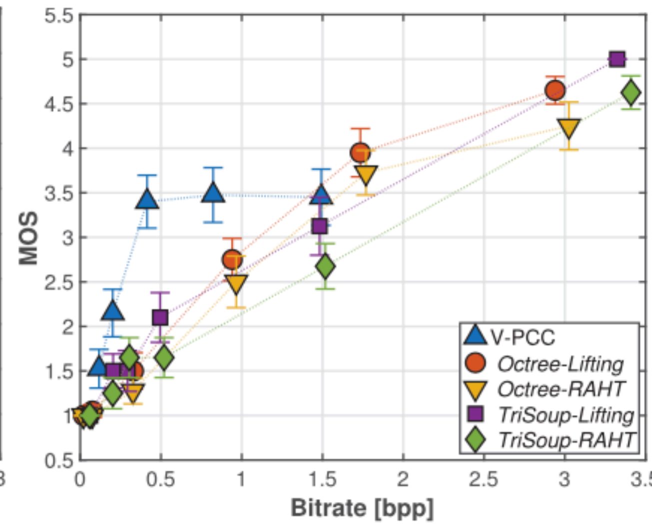
(e) *loot*



(f) *longdress*



(g) *soldier*



(h) *the20smaria*

Performance Analysis

- V-PCC outperforms all variants of G-PCC for low rates, especially for dense point clouds
  - V-PCC maintains, or even increase the number of output points while the quality is decreasing
- V-PCC doesn’t achieve not achieve transparent, or close to transparent quality
  - Saturation effect due to the presence of holes or occlusions, especially for sparse point clouds
- Regarding G-PCC, the following conclusions can be made:
  - Octree is equivalent or better than TriSoup, which introduces missing regions in the form of triangles
  - Lifting color module is marginally better than RAHT, which introduces blockiness due to the quantization of the transform coefficients

5.2 Objective Quality Assessment

Objective Evaluation

- Subjective tests are **time consuming, expensive, and difficult to design …**

- **Objective PC quality assessment metrics** aim to calculate scores with high level of correlation to the subjective scores and are used for
  - RD performance assessment of PC coding solutions
  - Optimization of PC coding solutions
  - Measuring end-to-end quality in PC streaming solutions, much more than coding

- **Point cloud quality metrics scope:**
  - Geometry or color only
  - Joint geometry-color
  - Static vs dynamic

Point-to-Point (Po2Point) Quality Metrics

- Po2Point quality metrics establish **correspondences** for each point in the original PC A and the nearest neighbor (NN) point in the degraded PC B
  - Correspondences are computed in two directions and, thus, also from PC B to PC A

$$d_{\mathbf{A,B}}^{\mathbf{Po2Point}}(i) = \left\| \vec{e}_1(a_i, b_j) \right\|_2^2$$

- **Mean Square Error (MSE):** average of the squared distance between each point and their corresponding nearest neighbor, for all points

$$\mathbf{MSE}_{\mathbf{A,B}} = \frac{1}{N_A} \sum_{\forall a_i \in A} d_{\mathbf{A,B}}^{\mathbf{Po2Point}}(i)$$

- **Geometric PSNR (MPEG D1):** MSE based including normalization factor P that depends on the PC coordinates precision (pr)

$$\mathbf{PSNR}_{\mathbf{A,B}} = 10\,\mathbf{log}_{10}\left(\frac{3P^2}{\mathbf{MSE}_{\mathbf{A,B}}}\right) \textbf{ with } P = 2^{pr} - 1$$

Point-to-Plane (Po2Plane) Quality Metrics

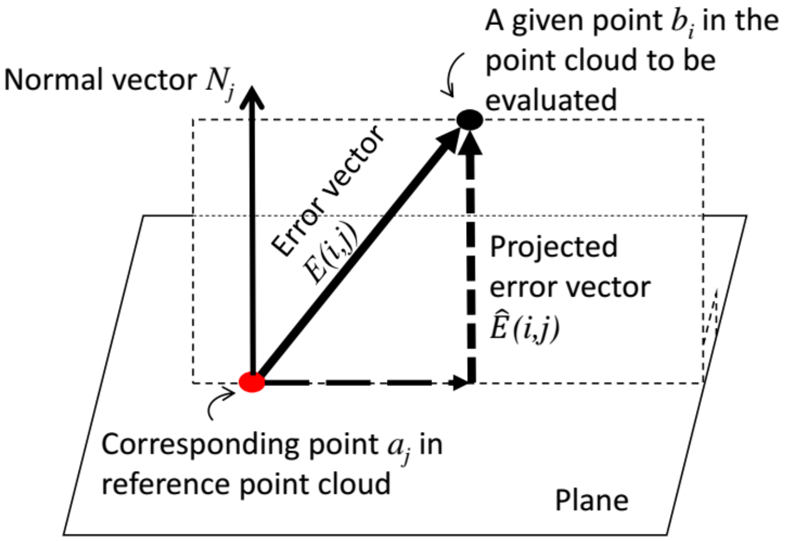
- **Po2Plane quality metrics** consider the points as part of **3D object surfaces**
  - Correspondences between points are computed as in Po2Point metrics

- Po2Point distance vector between these two points is projected on normal vector

$$d_{\mathbf{B,A}}^{\mathbf{Po2Plane}}(i) = \left\|\vec{e}_2(a_i, b_j)\right\|_2^2 = \left(\vec{e}_1\left(a_i, b_j\right) \cdot \overrightarrow{n_j^b}\right)^2$$

- Normals are computed in a neighbourhood around the selected point on the original point cloud

- **MSE distance and Geometric PSNR (MPEG D2)** are computed with the projected error distances as for Po2Point metrics



Dealing with Symmetry

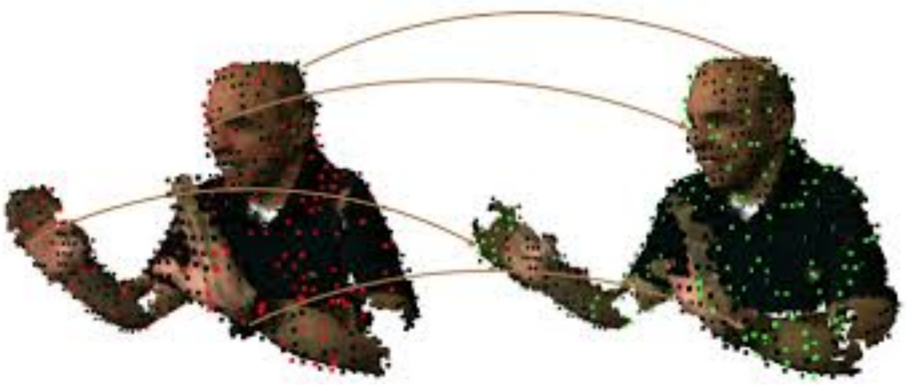
- Both metrics are calculated **symmetrically**, from the degraded point cloud to the reference point cloud and the opposite

Po2Point and Po2Plane

$$\text{MSE} = \max\left(\text{MSE}_{\mathbf{A,B}}, \text{MSE}_{\mathbf{B,A}}\right)$$

$$\text{HAUS} = \max\left(\text{HAUS}_{\mathbf{A,B}}, \text{HAUS}_{\mathbf{B,A}}\right)$$

$$\text{PSNR} = \min\left(\text{PSNR}_{\mathbf{A,B}}, \text{PSNR}_{\mathbf{B,A}}\right)$$

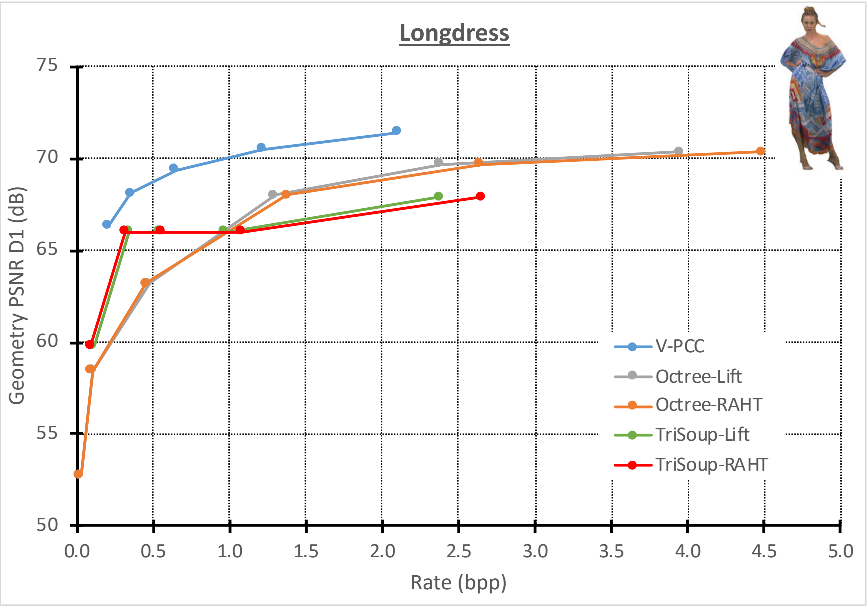


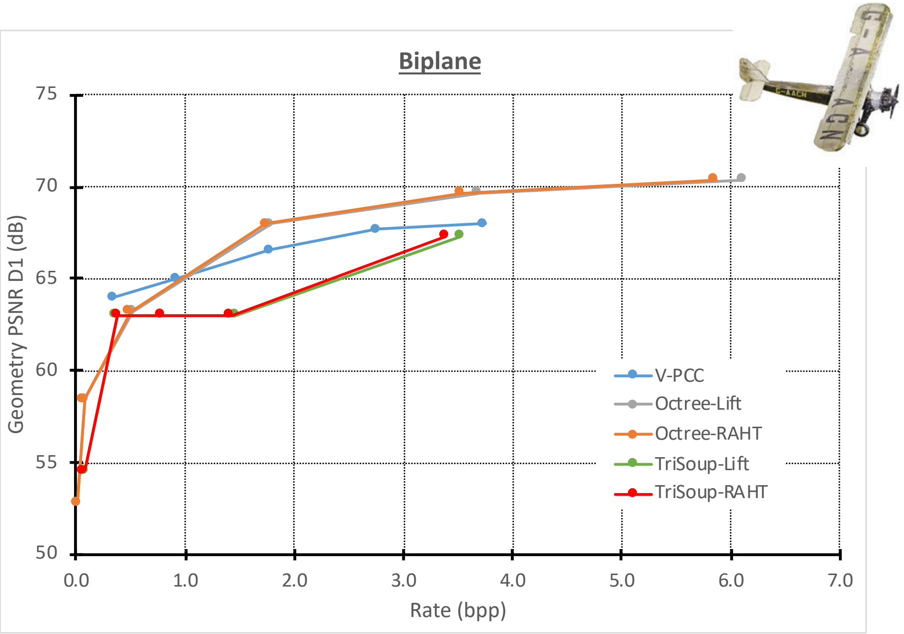
Extensions to Color and Time

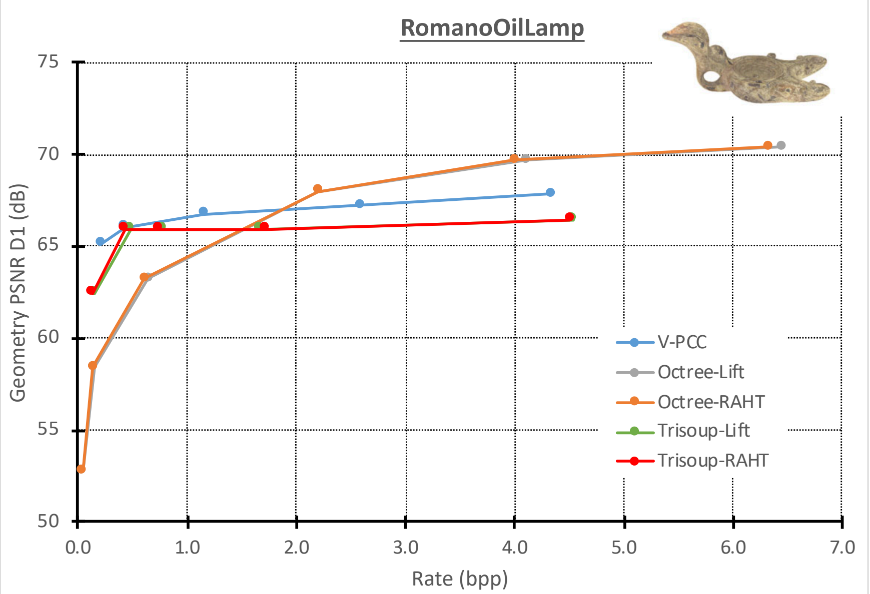
* Point cloud color quality metrics are mostly an **extension of geometry quality metrics** but applied to attributes (color)
    * Correspondence computation
    * RGB to YUV conversion is applied
    * Traditional pixel-wise 2D color quality metrics (e.g. PSNRY) are applied to the points involved in the correspondence

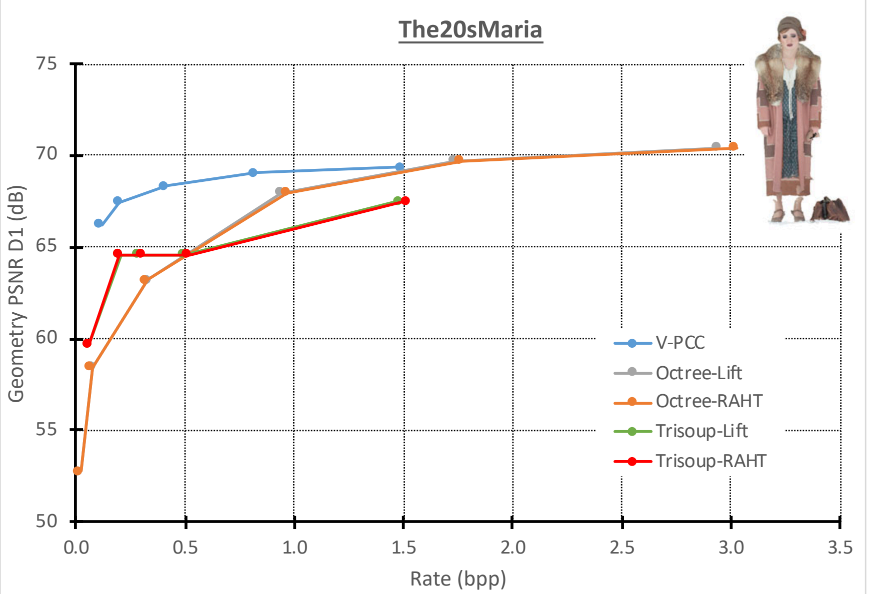
* Dynamic point cloud quality metrics are **mostly an extension of static quality metrics** with some temporal pooling (average)
    * Not much literature dealing with temporal aspects
    * Subjective evaluation for this scenario is much harder

RD Performance Po2Point Geometry PSNR

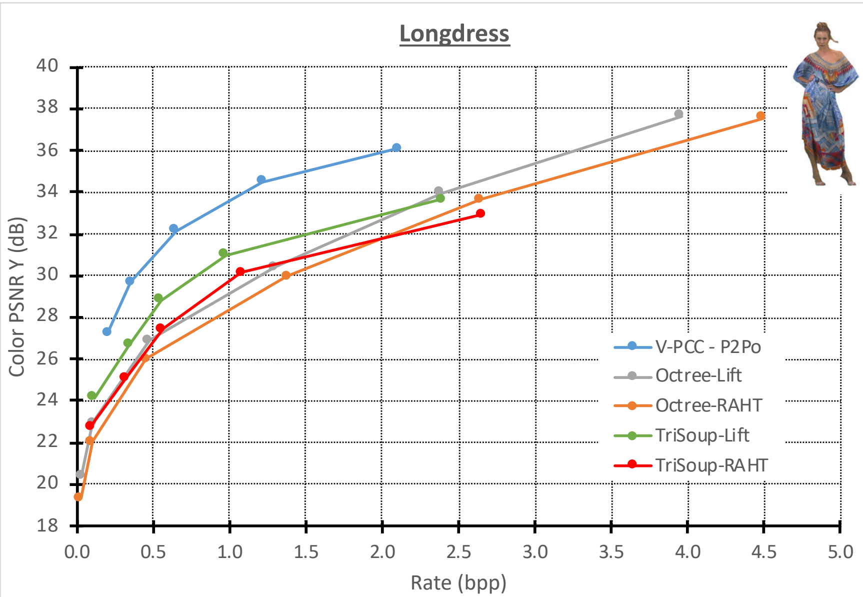


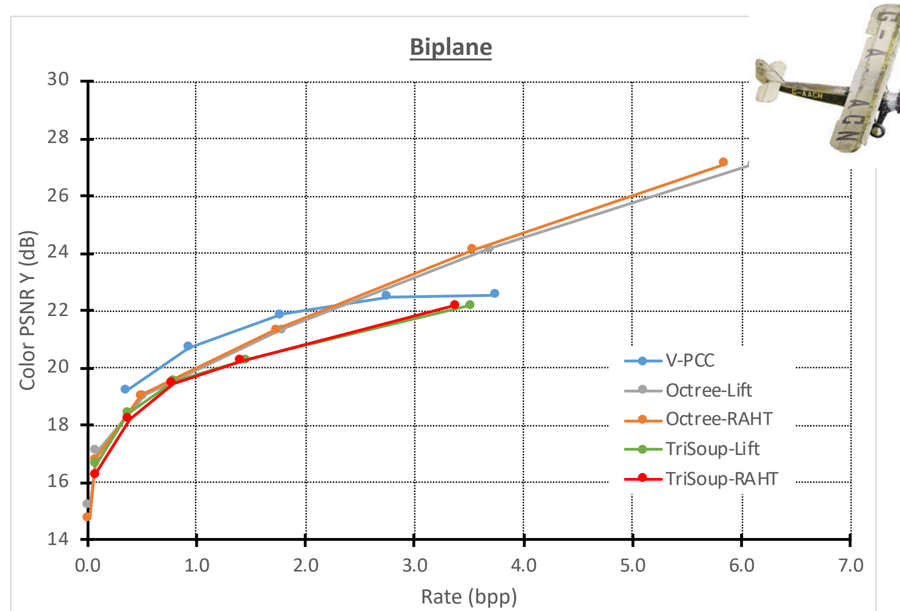


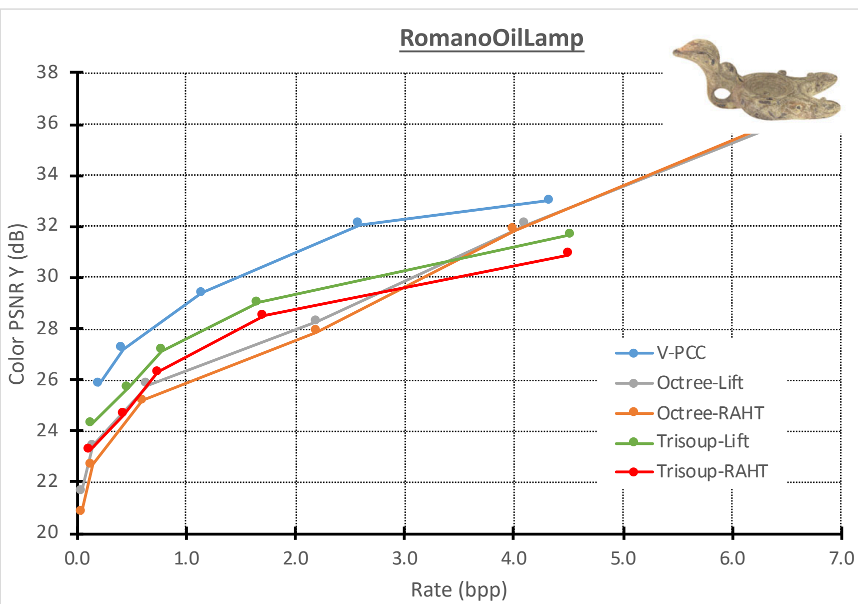


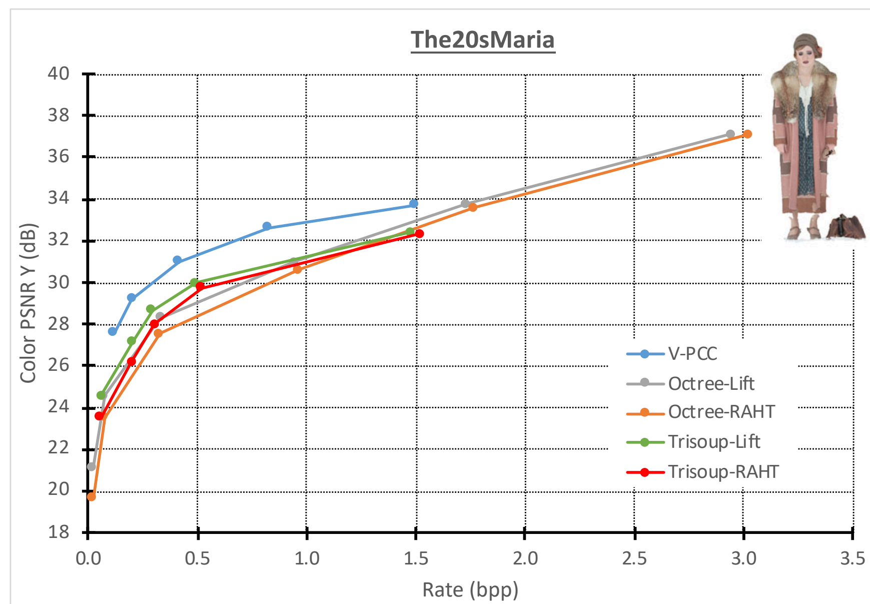


RD Performance Po2Point Color PSNR









6 Trends and Summary

Visual Data Coding: the Trends

- Since the 90s, **major video compression gains** have been obtained in an almost continuous way. However, this pace has been reducing …
- **Plenoptic imaging representation** should provide a major step forward towards **visual realism and immersion**
- **Novel capturing systems and immersive/interactive 3D viewing experiences** are creating new opportunities for future networks and technologies.
- For **new visual sensors/displays and new visual data representation models,** new, efficient coding solutions must be developed ☺
- Point cloud coding enables **interactive high quality 3D content** by providing manageable bitrates and also reducing requirements in creation, transmission and rendering of 3D content.
- **First Point Cloud Coding standards are ready (MPEG V-PCC & G-PCC) …**

The Take-Home Message

Point cloud visual representation is useful for many applications, notably to offer immersive experiences …

Since meaningful point clouds require many points and associated attributes, efficient point cloud coding solutions and standards are critical for the success of point cloud based applications !

TÉCNICO LISBOA

Thanks for your Attention!

And thanks to the colleagues who provided me many of the slides …


QUESTIONS
ANSWERS