**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2014/N16716**

 **Jan. 2017, Geneva, CH**

**Title: Draft Test Conditions and Complementary Test material**

**Source: MPEG 3DG**

**Status: Approved**

# Abstract

This document contains the test conditions and complementary test material for the Call for Proposals (CfP) for 3D point cloud compression technology, targeting an efficient representation of static objects and scenes, as well as dynamic objects and real-time acquisition environments. The document status is Draft, an updated version will be issued at MPEG Hobart meeting.

# Introduction

The Call for proposals for point cloud compression was issued after the 117th MPEG meeting in Geneva CH available as output document [1]. This additional document specifies the test conditions and test materials.

# Test Materials, Categories and Conditions

## Test Material Datasets

Below is a list of the 3D point cloud content sequences to be used, organized based on the test category. All datasets are available in the MPEG Content repository accessible under the following URL: <http://157.159.160.118/MPEG/PCC/DataSets/pointCloud/>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Category** | **Test Class** | **Filename** | **# Frms** | **# Pts** |
| Category 1. Static Objects and Scenes | Inanimate Objects | Egyptian\_mask | 1 |  |
| Landscape (00014) | 1 |  |
| Head (00039) | 1 |  |
| Frog (00067 ) | 1 |  |
| Shiva (00035) | 1 |  |
| Statue\_Klimt | 1 |  |
| Buildings and Facades | house without roof (57) | 1 |  |
| Arco Valentino Dense | 1 |  |
| Palazzo Carignano Dense | 1 |  |
| Façade9 (00009) | 1 |  |
| Façade15 (00015) | 1 |  |
| Façade64 (00064)  | 1 |  |
| People | Lincoln | 1 |  |
| Break Dancers | 1 |  |
| Large-scale static scenes | Stanford Area2 (all?) | 1 |  |
| Stanford Area3 (all?) | 1 |  |
| Category 2. Dynamic Objects | People | 8isequence 1 |  |  |
| 8isequence 2 |  |  |
| 8isequence 3 |  |  |
| 8isequence 4 |  |  |
| Technicolor sequence 1 |  |  |
| Lincoln |  |  |
| Break Dancers |  |  |
| Category 3. Dynamic Acquisition  | Frame-based laser scans | Ford Campus Vision |  |  |
| Urban Scene |  |  |
| Pomerleu |  |  |
| NCLT dataset |  |  |
| Canadian Planetary |  |  |

Notes:

1. The objects from Category 1,2 and 3 do not have normals or other attributes. The following information is available per point: X, Y, Z, R, G, B.
2. An additional dataset containing normals will be made available and this information will be used to compute the distortion metrics D2 point to plane
3. The order of the points is not necessary to be kept in decoded versions
4. For category 3 (dynamic acquisition) timestamps and point precision should be added in the .ply files
5. Lossless geometry, attribute, reflectance is important for category 3
6. The objects from category 2 do not have normals or other attributes. The following data is available per point: X, Y, Z, R, G, B, t

## Test Conditions and Parameters

The following terms are defined and used in the specification of the test conditions:

* **Lossy Geometry**: The decoded compressed content is not necessarily numerically identical to the uncompressed content. The number of points in the output cloud can be less than the number of points in the input cloud.
* **Lossless Geometry**: The decoded compressed content is numerically identical to the uncompressed content. The number of points in the output cloud is identical to the number of points in the input cloud
* **Lossy Attribute:** The decoded compressed attribute content is not necessarily numerically identical to the uncompressed content.
* **Lossless Attribute:** The decoded compressed attribute content is numerically identical to the uncompressed attribute content.
* **Spatial random Access:** It is possible to decode the point-cloud corresponding to a pre-defined region from the compressed point cloud.
* **Progressive/Scalable:** It is possible to first decode a coarse point cloud and then refine it with additional data from the compressed bit stream.
* **Intra-only**. Each frame can be decoded independently of any other encoded frames.
* **Temporal-random-access**. Point Clouds in a sequence can reference any other point cloud in the sequence for compression of the data. The entire sequence is encoded and decoded simultaneously in pre-defined groups of point clouds. .
* **Low-delay.** Frames in the sequence can be decoded with low-delay. This implies that the encoder and decoder do not need a reference to frames coded in the future.

### Test Conditions for Static Objects and Scenes

The static objects will be tested in two conditions:

* Lossy and lossless geometry with lossy attribute
* Spatial random access, progressive/scalable decoding

### Test Conditions for Dynamic Objects

Compression of dynamic point clouds will be tested under different conditions:

- Lossy geometry and lossy attributes

- Intra-only, temporal random access, low delay

For dynamic point clouds spatial random access and scalability/progressive will not be tested for responses to the CfP, however contributors are invited to describe if their solution can be enriched with such functionalities.

### Test Conditions for Dynamic Acquisition

Compression of dynamic acquisition will be tested under different conditions:

* Lossless (on geometry, reflectance and colors)
* Retain scan order (order of points within a frame)

A unified compression solution considering together positions, attributes and colors is suitable.

### Restrictions for all test categories and conditions

Point cloud compression technologies shall obey the following additional constraints:

1. Only use post-processing if it is part of the decoding process.
2. Optimization of encoding parameters using non-automatic means is discouraged.
3. The coding test set shall not be used as the training set for training large entropy coding tables, VQ codebooks, etc.
4. If any type of pre and post processing is used, it should be clearly described in the submission and anything that is not essential to the submission is strongly discouraged. If any preprocessing is done, the comparison shall be done with the original data.
5. Submitted bitstreams must have bitrates less or equal to those of the anchors

### Anchors

Anchors were generated by using the current software including the GOP structure (limited currently at 2 frames, however the call is for more complex GOP structures). Anchors will be generated by a GOP of 2 (IP), contributors should provide results for GOP of maximum 8 (IxxxxxxxI), with x being P or B. The software is based on the point cloud library 1.7.2 extended with features for lossy color coding, and bit-rate parameter settings and inter-predictive coding. The software is available publicly and in the mpeg svn under the links <https://github.com/RufaelDev/pcc-mp3dg/tree/trunk> and is available for MPEG members in the svn [http://wg11.sc29.org/svn/repos/MPEG-04/Part16-Animation\_Framework\_eXtension(AFX)/trunk/3Dgraphics/3DG-PCC/trunk/](http://wg11.sc29.org/svn/repos/MPEG-04/Part16-Animation_Framework_eXtension%28AFX%29/trunk/3Dgraphics/3DG-PCC/trunk/). The anchors are also available on http://mpegfs.int-evry.fr/MPEG/PCC/DataSets/pointCloud/CfP/anchors/

# Evaluation Procedure

##  Objective Evaluation

* D1: Point to Point Distance

The objective evaluation method (called D1) described in Annex C will be used as a full reference metric. D1 is a point to point distance. It will be used in each of the test conditions to provide a distortion metric. The metric compares the original data with the decoded data and provide a numerical value. Comparison will be made based on the bitrate and distortion targeting better performance compared to the corresponding rate points of the anchors.

Both PSNR vs MSE should be reported in the response excel sheet. Curves should be provided by using PSNR (normalized by using a constant pre-defined per sequence). To compute PSNR from MSE, a constant will be used.

For dynamic content D1 will be used and averaged over frames (average over I frames, average over P frames, average over all frames). In addition, there will be verification of the behavior of PSNR over time (ask contributors to provide PSNR per frame).

* D2: Point to Plane Distance (to be used for Cat1 and Cat2)

The computation of the normals (for sequences that don't have yet the normals) is done by PCC by using 12 nearest neighbors.

* Other attributes distortion

For the colors the distortion is measured in YUV (3 distortions) with the weight of 6:1:1 in PSNR (Y:U:V). By default Y, U, V PSNRs will be individually used to compare methods. Inform contributors that weighted PSNR will be used if needed. Metric available in experimental software platform. Base consideration: color distortion measurements will be similar as for video.

For the reflectance/reflectivity (I), the distortion measured as with Y component from color attributes.

* **Other axes of evaluation:**
1. Coverage of various bitrate ranges
2. Complexity vs performance

A complete complexity analysis for encoder and decoder should not be provided by contributors but compression/decompression time should be reported per sequence (a relative time with respect to ALL anchors from category they target for encode/decode process should be provided).

1. LOD and spatial random access points (SRAP) support (Cat1, Cat2, Cat3)

Proponents should describe how LOD/scalability and SRAP can be added to their solution (if is not already the case).

##  Subjective Evaluation

The subjective evaluation will be done based on expert viewing. The point clouds will be rendered using a simple points rendering system such as provided with the MPEG-PCC-Exploration Software or other rendering methods specified. The rendering view-point will follow a pre-defined path. The output will be stored as high quality video sequences of a clearly defined length as close as possible to 10s. These video sequences will be viewed by different experts and scored on a 1 to 5 scale (reference the ITU document describing the procedure – ITU-R BT.500).

**Notes on the rendering**

Some renderer will be used for generated all the sequences. The final renderer will be selected at the MPEG Hobart meting.

# References

1. Call for proposals for point cloud compression ISO/IEC JTC1/SC29 WG11 n16732, Geneva CH, January 2017
2. Use Cases for Point Cloud Compression, ISO/IEC JTC1/SC29 WG11 Doc. N16331, Geneva, CH, June 2016.
3. Requirements for Point Cloud Compression, ISO/IEC JTC1/SC29 WG11 Doc. N16330, Geneva, CH, June 2016.

#

# Annex A: Evaluation Sheet (to be filled during evaluation phase/also to be used for self-evaluation)

Excel sheet – to be provided at the Hobart meeting

# Annex A: Objective Evaluation Method D1

The point cloud consists of a set of points represented by (x,y,z) and various attributes of which color components (y,u,v) are of critical importance.

For a detailed description of the point cloud representation as defined in MPEG we refer to Annex D.

First, let us define the point *v,* it has as a mandatory position in a 3D space *(x,y,z)* and an optional color attribute *c*  that has components *r,g,b* or *y,u,v* and optional other attributes possibly representing normal or texture mappings.

$point v=\left(\left((x,y,z\right),\left[c\right],\left[a\_{0}..a\_{A}\right]\right): x,y,z\in R ,\left[c \in \left(r,g,b\right) \right|r,g,b \in N], [a\_{i} \in \left[0,1\right]])$ (def. 1)

The point cloud is then simply a set of K points without a strict ordering:

$Original Point Cloud V\_{or}=\{\left(v\_{i}\right):i=0…..K-1\}$ (def. 2)

The point cloud is a set of (x,y,z), and attributes can be attached to the points. The original cloud Vor will act as the reference for determining the quality of a second degraded cloud Vdeg. Vdeg consists of N points, where not necessarily N = K. Vdeg is a version with a lower quality possibly resulting from lossy encoding and decoding of Vor. This can result in a different point count N.

$Degraded Point Cloud V\_{deg}=\{\left(v\_{i}\right):i=0…..N-1\}$ (def. 3)

The quality metric $Q\_{point\\_cloud}$ is computed from $V\_{or}$ and $V\_{deg}$ and used for assessment as shown in Figure 1.



Figure 1 Schematic view of full reference quality estimation of point cloud codec

In Table 1 we outline the metrics used for the assessment of the quality of a point cloud. The geometric distortion metrics are similar as the standard ones for meshes based on haussdorf (Linf) and root mean square (L2), instead of distance to surface we take the distance to the most nearby point in the cloud (see defs 4,5,6,7). We defined PSNR as the peak signal of the geometry over the symmetric rms distortion (def 8.). For colors we defined a similar metric, we compare the color of the original cloud to the most nearby colour in the degraded cloud and compute PSNR per YUV component in the YUV color space (def. 10). The key advantage of this metric is that it corresponds to PSNR in Video Coding. The quality metric is supported in the 3DG PCC software [3].

Table 1 Assessment criteria for assessment of the point cloud quality of Vdeg, $Q\_{point\\_cloud}$

|  |  |
| --- | --- |
| d\_symmetric\_rms | Symmetric rms distance between the point clouds (def. 5.) |
| d\_symmetric\_haussdorf | Symmetric haussdorf distance between the clouds (def. 7.) |
| psnr\_geom | Peak signal to noise ratio geometry (vertex positions) (def. 8.) |
| psnr\_y | Peak signal to noise ratio geometry (colors Y) (def. 10) |
| psnr\_u | Peak signal to noise ratio geometry (colors U) (as def. 10 rep. y for u) |
| psnr\_v | Peak signal to noise ratio geometry (colors V) (as def. 10 rep. y for v) |

$d\_{rms}\left(V\_{or},V\_{deg}\right)=\sqrt{\frac{1}{K}\sum\_{vo\in Vor}^{}\left⟦vo-vd\\_nearest\\_neighbour\right⟧^{2}}$ (def.4)

$d\_{symmetric\\_rms}\left(V\_{or},V\_{deg}\right)=max⁡(d\_{rms}\left(V\_{or},V\_{deg}\right),d\_{rms}\left(V\_{deg},V\_{or})\right)$ (def.5)

$d\_{haussdorf}\left(V\_{or},V\_{deg}\right)= max\_{v\_{o} \in V\_{or, }}( ||v\_{o}-v\_{d\\_nearest\\_neighbour}||\_{2} , v\_{d} is the point in Vdeg closest to v\_{o} (L2)$) (def.6)

$d\_{symmetric\\_haussdorf}\left(V\_{or},V\_{deg}\right)=max⁡(d\_{haussdorf}\left(V\_{or},V\_{deg}\right),d\_{haussdorf}\left(V\_{deg},V\_{or}\right)$ (def. 7)

$BBwidth=max⁡(\left(xmax-xmin\right),\left(ymax-ymin\right),(zmax-zmin)$ (def.8)

$psnr\_{geom}=10log\_{10}(\left|BBwidth\right||\_{2}^{2}/ (d\_{symmetric rms}\left(V\right))^{2})$ (def. 9)

$d\_{y}\left(V\_{or},V\_{deg}\right)=\sqrt{\frac{1}{K}\sum\_{vo\in Vor}^{}\left⟦y(vo)-y(v\_{dnearest\_{neighbour}})\right⟧^{2}}$ (def.10)

$psnr\_{y}=10log\_{10}(\left|255\right||\_{}^{2}/ (d\_{y}\left(V\_{or},V\_{deg}\right)^{2})$ (def. 11)

Additional metrics that define the performance of a codec are outlined in Table 2. Geometry and color byte size are mandatory (if colors are available). Encoding and decoding times are optional as they can only be an indicator of complexity. The number of input and output points shall also be reported.

Table 2 Assessment criteria for assessment of the performance of point cloud compression ratio

|  |  |
| --- | --- |
| Compressed size | Complete compressed mesh size |
| In point count | K, the number of vertices in Vor |
| Out point count | N, number of vertices in Vdeg |
| Bytes\_geometry\_layer | Number of bytes for encoding the vertex positions |
| Bytes\_color\_layer (opt) | Number of bytes for encoding the colour attributes |
| Bytes\_att\_layer (opt) | Number of bytes for encoding the other attributes |
| Encoder time (opt) | Encoder time in ms on commodity hardware (optional) |
| Decoder time (opt) | Decoder time in ms on commodity hardware (optional) |

All metric shall be supported by the MPEG exploration software evaluation benchmark.

# Annex B: MPEG Point Cloud and PCC Definitions

*Point Cloud*. A point cloud is defined as set of (x,y,z) coordinates, where x, y, z have finite precision and dynamic range. Each (x,y,z) can have multiple attributes associated to it (a1 ,a2, a3 …). Typically each point in a cloud has the same number of attributes attached to it. Point clouds with different number of attributes per point will not be considered unless explicitly proven to be useful.

$point v=\left(\left((x,y,z\right),\left[c\right],\left[a\_{0}..a\_{A}\right]\right): x,y,z\in R ,\left[c \in \left(r,g,b\right) \right|r,g,b \in N], [a\_{i} \in \left[0,1\right]])$ (def. 1)

The point cloud is then simply a set of K points without a strict ordering:

$Original Point Cloud V\_{or}=\{\left(v\_{i}\right):i=0…..K-1\}$ (def. 2)

*Lossless Point Cloud Compression*. In the case of lossless compression, the decoder returns exactly the same set of (x,y,z), with exactly the same attributes. This is the same number of points with the same coordinates. An efficient way to canonically order the set and test for equality is to convert to Morton codes [14].

*Lossy Point Cloud Compression*. In this case the number of points in the set and/or the positions x,y,z are not identical to the original.

*Lossy Attributes Compression*. In this case the values of the attributes are not the same compared to the values of the original