TÉCNICO LISBOA

**UNIVERSIDADE DE LISBOA**

**INSTITUTO SUPERIOR TÉCNICO**

# Point Cloud Quality Assessment

Alireza Javaheri

**Supervisor:** Doctor João Miguel Duarte Ascenso

**Co-supervisor:** Doctor Fernando Manuel Bernardo Pereira

Thesis approved in public session to obtain the PhD Degree in

Electrical and Computer Engineering

Jury final classification: Pass with Distinction

2021

Obra entregue para apreciação da prova de Doutoramento em Engenharia Eletrotécnica e de Computadores para obtenção do grau de Doutor concedido pela Universidade de Lisboa, no Instituto Superior Técnico, ao abrigo do D.L. 74/2006/, de 24 de março.

Esta obra não poderá ser reproduzida sem expresso consentimento do autor, ressalvadas as disposições constantes do art.º 76.º do código do direito do autor (Decreto-lei n.º 63/85, de 14 de março).

Lisboa, 1 de julho de 2021

O presidente do júri

Homologo

**UNIVERSIDADE DE LISBOA**

**INSTITUTO SUPERIOR TÉCNICO**

# Point Cloud Quality Assessment

Alireza Javaheri

**Supervisor:** Doctor João Miguel Duarte Ascenso

**Co-supervisor:** Doctor Fernando Manuel Bernardo Pereira

Thesis approved in public session to obtain the PhD Degree
in Electrical and Computer Engineering
Jury final classification: Pass with Distinction

Jury

Chairperson: Doctor Mário Alexandre Teles de Figueiredo, Instituto Superior Técnico, Universidade de Lisboa

Members of the Committee:

Doctor Touraj Ebrahimi, School of Engineering, École Polytechnique Fédérale de Lausanne, Suíça
Doctor Paulo Luís Serras Lobato Correia, Instituto Superior Técnico, Universidade de Lisboa
Doctor Maria Paula dos Santos Queluz Rodrigues, Instituto Superior Técnico, Universidade de Lisboa
Doctor João Miguel Duarte Ascenso, Instituto Superior Técnico, Universidade de Lisboa
Doctor Nuno Miguel Morais Rodrigues, Escola Superior de Tecnologia e Gestão, Politécnico de Leiria

Funding Institutions

This research has been made possible with funding from the Fundação para a Ciência e a Tecnologia, Instituto de Telecomunicações.

2021

*To My wonderful wife,*

***Faezeh***

*who was always beside me in this chapter of life.*

# Abstract

Nowadays, richer 3D visual representation formats are emerging, notably light fields and point clouds. These formats enable new applications in many usage domains, notably virtual and augmented reality, geographical information systems, immersive communications, and cultural heritage. Recently, following major improvements in 3D visual data acquisition, there is an increasing interest in point-based visual representation, which models real-world objects as a cloud of sampled points on their surfaces. Point cloud is a 3D representation model where the real visual world is represented by a set of 3D coordinates (the geometry) over the objects with some additional attributes such as color and normals. With the advances in 3D acquisition systems, it is now possible to capture a realistic point cloud to represent a visual scene with a very high resolution. These point clouds may have up to billions of points and, thus, storing and transmitting them in a raw format would require an unbearable amount of memory and bandwidth. Therefore, the storage and transmission of large point clouds critically ask for the development of efficient point cloud coding solutions. In this context, to boost a wide adoption of this 3D visual representation model, it is also necessary to reliably measure the quality of experience offered to the end-users by measuring the point cloud quality. While objective quality assessment metrics aim to mathematically measure the quality of point clouds, notably decoded point clouds, ideally replicating the scores that would be given by human beings, the subjective quality assessment allows not only to perform more reliable assessment but also allows to assess the correlation of the available objective quality metrics with the users' opinion scores. The design and identification of the most reliable objective quality metrics, notably for point clouds, requires subjective evaluation data obtained in meaningful and already proven methodology.

In this context, the main objective of this Thesis is twofold: first, to perform appropriately designed subjective quality assessment experiments for decoded point clouds under different degradations and impacting factors like coding and rendering which allows to assess and benchmark the reliability of point cloud objective quality metrics; and second, to propose novel objective quality metrics with a higher correlation with the obtained subjective assessment scores.

To achieve these objectives, three subjective quality assessment experiments have been performed considering different contents, degradations, and impact factors. Moreover, four objective quality metrics have been proposed, all outperforming the state-of-the-art objective quality metrics at the time they were developed. Due to the importance of geometry on point cloud perceived quality, and new challenges associated with geometry quality evaluation, most of the efforts in this Thesis were around geometry quality evaluation, notably for static point clouds. However, in the last chapter, a quality metric jointly considering geometry and color is proposed, which outperforms all available quality metrics in the literature for point clouds.

**Keywords:** *Point Cloud, Subjective Quality Assessment, Objective Quality Assessment, Geometry, Attributes, Coding, Rendering.*

# Resumo

Hoje em dia, formatos mais ricos de representação visual 3D estão a surgir, nomeadamente campos de luz e nuvens de pontos. Estes formatos tornam possível oferecer novas aplicações em vários domínios, nomeadamente realidade virtual e aumentada, sistemas de informação geográfica, comunicações imersivas e aplicações culturais. Recentemente, no seguimento de vários avanços em termos da aquisição de informação visual 3D, cresceu o interesse em representações visuais que modelam os objectos do mundo real como uma nuvem de pontos amostrados sobre as suas superfícies. As nuvens de pontos são um modelo de representação 3D onde o mundo visual real é representado através de um conjunto de coordenadas 3D sobre os objectos (a geometria) com alguns atributos adicionais tais como cor e normais. Com os avanços nos sistemas de aquisição 3D, é agora possível capturar nuvens de pontos realistas para representar uma cena visual com elevada resolução. Estas nuvens de pontos podem ter até alguns biliões de pontos e logo o seu armazenamento e transmissão no formato original requereria uma quantidade irrazoável de memória e débito de transmissão. Assim sendo, o armazenamento e transmissão de nuvens de pontos exige o desenvolvimento de soluções eficientes de codificação.

Neste contexto, de forma a possibilitar a adopção em larga escala deste modelo de representação visual 3D, é também necessário medir de um modo fiável a qualidade de experiência oferecida aos utentes, medindo a qualidade das nuvens de pontos. Enquanto as métricas objectivas de avaliação de qualidade têm como finalidade medir a qualidade das nuvens de pontos de forma matemática, nomeadamente de nuvens de pontos descodificadas, idealmente replicando as classificações que seriam dadas por observadores humanos, a avaliação subjectiva de qualidade permite não só fazer uma avaliação mais fiável da qualidade mas também disponibilizar os dados necessários para avaliar a correlação das classificações objectivas de qualidade com as classificações subjectivas. O desenvolvimento e a identificação das métricas objetivas de qualidade mais fiáveis, especialmente para nuvens de pontos, necessita de dados de avaliação subjetivos obtidos com uma metodologia comprovada.

Neste contexto, esta Tese tem dois objectivos principais: primeiro, realizar experiências de avaliação da qualidade subjectiva de nuvens de pontos descodificadas, com diferentes tipos de degradação e factores de impacto como a codificação e a renderização, o que permite avaliar e comparar a fiabilidade das métricas objectivas de avaliação de qualidade já disponíveis para nuvens de pontos; segundo, propor novas métricas objectivas de avaliação de qualidade para nuvens de pontos, nomeadamente com uma melhor correlação com as correspondentes avaliações subjectivas.

Para alcançar estes objectivos, foram realizadas três experiências de avaliação subjectiva de qualidade considerando diferentes conteúdos, degradações e factores de impacto. Para além disso, foram propostas quatro métricas objectivas de avaliação de qualidade para nuvens de pontos, todas elas oferecendo um desempenho de correlação superior, às métricas de qualidade disponíveis no momento em que foram desenvolvidas. Devido à importância da geometria na qualidade perceptual das nuvens de pontos, a maioria do trabalho desenvolvido nesta Tese está focado na avaliação da qualidade da geometria, nomeadamente para nuvens de pontos estáticas. Contudo, no

último capítulo, é proposta uma métrica objectiva de qualidade que considera conjuntamente a geometria e a cor, apresentando um desempenho superior a todas as métricas objectivas de qualidade disponíveis na literatura para nuvens de pontos.

**Palavras-chave**: *Nuvem de Pontos, Avaliação Subjectiva de Qualidade, Avaliação Objectiva de Qualidade, Geometria, Atributos, Codificação, Renderização.*

# Acknowledgments

Accomplishing this PhD Thesis would not have been possible without the energy and support that I received from many people.

Foremost, I wish to express my deepest gratitude to my Ph.D. supervisors, Prof. João Ascenso and Prof. Fernando Pereira, for their incredible supervision and constant support. They convincingly guided and encouraged me to be professional and do the right thing even when the path got tough. I was so lucky for having them as my PhD mentors.

My sincere thanks also go to Prof. Catarina Brites. She was an experienced and trusted adviser for me. Her persistent help and encouragement were a milestone in the completion of this Thesis.

I am grateful to the members of the committee, Prof. Touraj Ebrahimi, Prof. Nuno Rodrigues, Prof. Paulo Correia, Prof. Paula Queluz, and Prof. Mário Figueiredo for the comprehensive analysis of the Thesis and their insightful and remarkable comments during the presentation that made this document improve.

I am whole-heartedly grateful to my colleagues and friends at Instituto de Telecomunicações for all the unforgettable memories we created together during these years, especially to be named Alireza Sepas-moghadam, André Guarda, Falah Jabar, Ivo Sousa, Antoine Dricot, Miguel Simões, and Tanmay Verlekar.

Words cannot express my love and gratitude to my mother, Parvin, my father, Mohammadmahdi, and my siblings, Marziyeh and Amir for all the sacrifices they have unconditionally made for me.

Most significantly, I wish to thank my beloved wife, Faezeh, for all her unconditional continual love, encouragement, and support from the bottom of my heart. It was with her emotional support that I could pass this state of my life.

In Addition, I am thankful to all the IT staff for the assistance and facilities, they provided me to conduct my PhD research.

This work would not have been possible without the financial support of Instituto de Telecomunicações (IT), Instituto Superior Técnico (IST), Fundação para a Ciência e a Tecnologia (FCT).

Alireza Javaheri

Lisbon, February 2021

# Table of Contents

# List of Figures

# List of Tables

# List of Acronyms

| | |
|---|---|
| 2D | 2 Dimensional |
| 3D | 3 Dimensional |
| ACR | Absolute Category Rating |
| ADMM | Alternating Direction Method of Multipliers |
| ANN | Average Nearest Neighbor |
| APD | Average Planar Distance |
| AR | Augmented Reality |
| CTC | Common Test Conditions |
| DCT | Discrete Cosine Transform |
| DISTS | Deep Image Structure and Texture Similarity |
| DoF | Degree of Freedom |
| DSIS | Double Stimulus Impairment Scale |
| FR | Full Reference |
| FSIM | Feature Similarity Index |
| GCM | Gaussian Color Model |
| GH | Generalized Hausdorff |
| GIS | Geographical Information System |
| G-PCC | Geometry-based Point Cloud Compression |
| GSP | Graph Signal Processing |
| GSPBox | Graph Signal Processing Toolbox |
| GT | Graph Transform |

| | |
|---|---|
| HaarPSI | Haar Perceptual Similarity Index |
| HAUS | Hausdorff Distance |
| HD | High Definition |
| HEVC | High Efficiency Video Coding |
| HMD | Head Mounted Device |
| HR | Hidden Reference |
| HR | Hidden Reference |
| ICIP | International Conference on Image Processing |
| IRPC | Instituto Superior Técnico Rendered Point Cloud |
| IST | Instituto Superior Técnico |
| IT | Instituto de Telecomunicações |
| ITU-R | International Telecommunication Union - Radiocommunication Sector |
| ITU-T | International Telecommunication Union - Telecommunication Sector |
| JS | Jensen Shannon |
| LBP | Local Binary Pattern |
| LLP | Local Luminance Pattern |
| LoD | Level of Details |
| LPIPS | Learned Perceptual Image Patch Similarity |
| MAE | Mean Absolute Error |
| MCCV | Monte Carlo Cross Validation |
| MMD | Mean Mahalanobis Distance |
| MMSP | MultiMedia Signal Processing |
| MNN | Maximum Nearest Neighbor |

| | |
|---|---|
| MOS | Mean Opinion Score |
| M-PCCD | MPEG Point Cloud Compression Dataset |
| MPEG | Moving Picture Experts Group |
| MSE | Mean Squared Error |
| MSMD | Mean Squared Mahalanobis Distance |
| MS-SSIM | Multi-Scale Structural Similarity Index Metric |
| NN | Nearest Neighbor |
| NR | No Reference |
| P2D | Point-to-Distribution |
| PC | Point Cloud |
| PCL | Point Cloud Library |
| PCQM | Point Cloud Quality Metric |
| Pl2Pl | Plane-to-Plane |
| PLCC | Pearson Linear Correlation Coefficient |
| Po2Pl | Point-to-Plane |
| Po2Po | Point-to-Point |
| PSNR | Peak Signal-to-Noise Ratio |
| PWC | Pairwise Comparison |
| QoE | Quality of Experience |
| QoMEX | Quality of Multimedia Experience |
| QP | Quantization Parameter |
| RA | Resolution-Adaptive |
| RAHT | Region Adaptive Hierarchical Transform |

| | |
|---|---|
| RD | Rate-Distortion |
| RMSE | Root Mean Squared Error |
| RR | Reduced Reference |
| SROCC | Spearman Rank Order Correlation Coefficient |
| SSIM | Structural Similarity Index Metric |
| TK | Tikhonov |
| TV | Total Variation |
| VIFP | Visual Information Fidelity Measure |
| V-PCC | Video-based Point Cloud Compression |
| VR | Virtual Reality |
| VSI | Visual Saliency Index |
| VV | Volumetric Video |

# Part I. Background

# Chapter 1

# Introduction

## 1.1 Context and Motivation

Nowadays, the emergence of new 3D visual representation models allows for more immersive experiences compared to the classical 2D images and videos. Moreover, recent advances in 3D acquisition have made these representation models increasingly popular. Point clouds are one of these immersive 3D representation models and allow offering many new visual-based applications such as geographical information systems, virtual and augmented reality, cultural heritage, and free viewpoint broadcasting. In this context, point clouds are becoming an important 3D visual representation format to capture the real visual world, easily and precisely, due to the availability of several novel acquisition devices, from range sensors to multi-camera arrays.

A point cloud (PC) is an unordered set of 3D points represented by their (x, y, z) coordinates and associated attributes, such as color, normal and reflectance. Figure 1.1 shows an example of a point cloud with point coordinates and attributes. Point clouds can be classified with respect to their temporal evolution as static, dynamic, or progressive point clouds. While static point clouds correspond to a single time instant, dynamic point clouds correspond to a point cloud evolving along time, thus corresponding to a sequence of static point cloud frames. Moreover, progressive point clouds correspond to large-scale point clouds that are not consumed all at once and, thus, are made from complementary parts of a visual scene; these parts are basically complementary static point clouds which may even have been acquired at different times; however, unlike dynamic point clouds, there is no temporal redundancy between these parts. These point clouds are often used in autonomous driving.

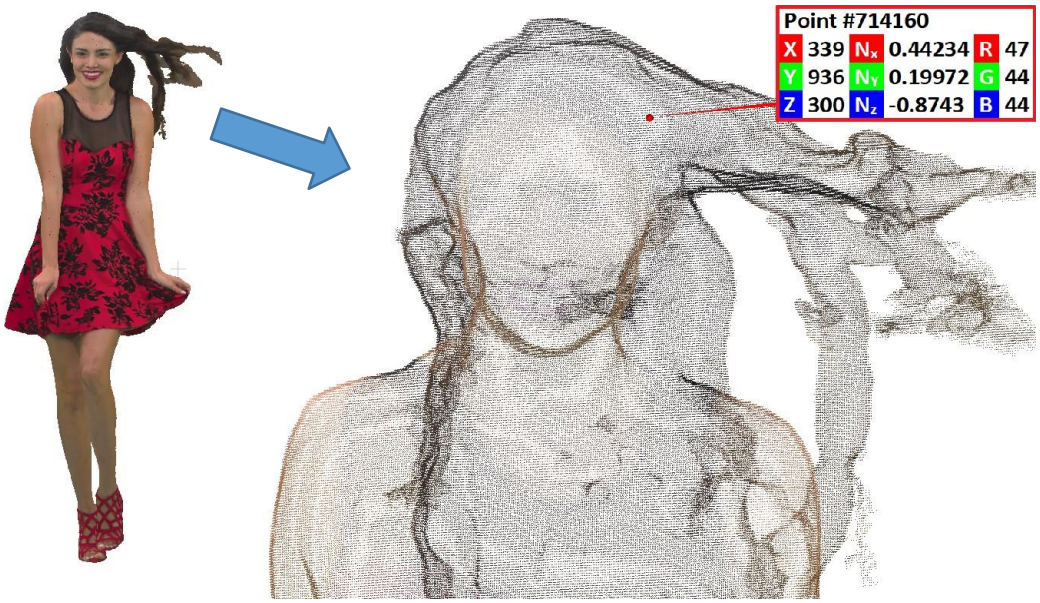



**Figure 1.1:** *Red and black* PC: full point cloud, and zoomed region, and a point specified with its (*x*, *y*, *z*) coordinates, normal coordinates, and associated R, G, B color.

The point cloud visual representation model is specifically relevant for multiple use cases such as immersive telecommunications, 3D sport replays broadcasting, Geographic Information Systems (GIS), cultural heritage, etc. [1]. Figure 1.2 shows some examples of point cloud applications.

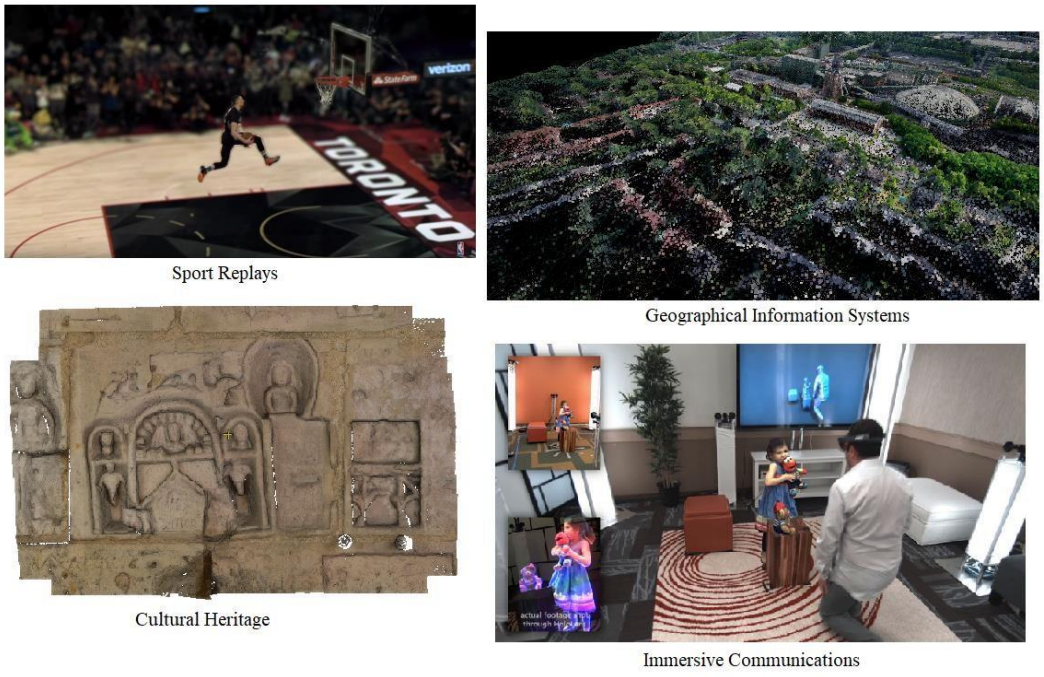



**Figure 1.2:** Point clouds associated to different applications such as sports replays, geographical information systems, cultural heritage, and immersive communications.

The major challenge associated to the point cloud representation model is the huge amount of data required for a visual scene with high fidelity, which may require several millions or even billions of points. Since this large amount of data needs to be efficiently stored and transmitted, point cloud coding solutions are essential to allow the deployment of point cloud-based applications involving the huge amounts of point cloud data produced by the point cloud acquisition devices and processes. As in the past for images and videos, the point cloud coding solutions must be efficient in the sense that a target fidelity/quality must be achieved at a much reduce rate compared to the raw, acquisition rate. To assess the rate-distortion (RD) performance of a point cloud coding solution, reliable point cloud quality assessment is required, notably to evaluate the fidelity/quality of the decoded point clouds for a specific rate. The reference way to reliably measure point cloud quality is through subjective quality assessment experiments where opinion scores are collected from several observers in a specially designed subjective testing framework. However, subjective quality assessment is expensive and time-consuming, and thus, reliable objective quality metrics are critical to effectively design efficient point cloud coding solutions. A reliable full-reference quality assessment metric can measure the quality of a decoded point cloud in comparison to the corresponding reference point cloud. Subjective quality experiments are performed not only to assess the point cloud quality more reliably, e.g., while developing a new codec, but also to provide ground truth subjective scores to assess the correlation performance of objective quality metrics with the quality perceived by humans.

The most used point cloud objective quality metrics are the so-called D1-PSNR and D2-PSNR metrics, widely used by MPEG [2], notably for the development of the emerging MPEG point cloud Coding (PCC) standards, the Geometry-based Point Cloud Compression (G-PCC) [3] and the Video-based Point Cloud Compression (V-PCC) [4] standards. Recent advances in point cloud coding, notably adopting different coding approaches and, thus, involving different coding artefacts, including deep learning-based point cloud coding, have made the point cloud quality assessment task more challenging. Several objective quality metrics have been proposed to address these challenges and several subjective quality assessment studies in the literature have assessed the correlation of available objective quality metrics for different coding and rendering conditions. Although major advances have been made, all these efforts show that there is still much room to improve objective quality metrics for point clouds

## 1.2 Objectives

This Thesis focuses on subjective and objective point cloud quality assessment, notably studying the impact of several modules in the point cloud processing pipeline on the perceived quality, such as the coding and rendering processes, and proposing more reliable point cloud objective quality metrics. The main motivations for the work developed in this Thesis were:

- There were not enough subjective quality assessment studies for point cloud quality, notably considering different contents, rendering and degradations.

- Most of the subjective quality assessment studies in the literature were considering pure octree pruning and Gaussian noise addition to create the test materials and not advanced coding solutions, such as the recent MPEG point cloud coding standards [5] [6] [7] [8].
- Objective quality metrics were not showing high correlation performance with subjective scores for the emerging point cloud coding solutions.

In this context, the key objectives of this Thesis were:

1. Study the impact of different types of distortions, notably coding artifacts, on the point cloud perceived quality.
2. Study the impact of different rendering solutions on the perceived quality of point clouds.
3. Benchmark the correlation performance of the state-of-the-art objective quality metrics for point clouds coded with state-of-the-art point cloud codecs for different coding conditions and rendering scenarios.
4. Design more reliable objective quality metrics for point clouds, notably with a better correlation with the perceived quality as measured by Mean Opinion Scores (MOS) scores, considering various point cloud intrinsic characteristics as well as point cloud coding and rendering solutions

## 1.3 Contributions

The research developed in the context of this Thesis has been focusing on the evaluation of the available objective quality metrics and the proposal of novel metrics to assess the quality of decoded point clouds. Naturally, new metrics should offer better correlation with subjective scores as expressed through Mean Opinion Score (MOS) and, thus, several subjective assessment experiments were performed, which by themselves are also an important contribution.

The subjective experiments and the objective quality metrics have considered different types of quality degradations resulting from different processes, notably coding, and rendering. Moreover, the subjective quality studies were performed considering different relevant factors that may impact the visibility of the degradations, such as the intrinsic point cloud characteristics, e.g., density.

Following the main objectives defined above, the main contributions of this Thesis are illustrated in Figure 1.3. The contributions are organized based on the two main objectives, i.e., subjective quality assessment studies and novel objective quality assessment metrics. Although the focus of this Thesis is on coding artefacts, denoising algorithms were also studied. The novel contributions made will be presented in detail in Part II (Chapters 3, 4 and 5) and Part III (Chapters 6, 7, 8 and 9), for subjective and objective quality assessment, respectively.

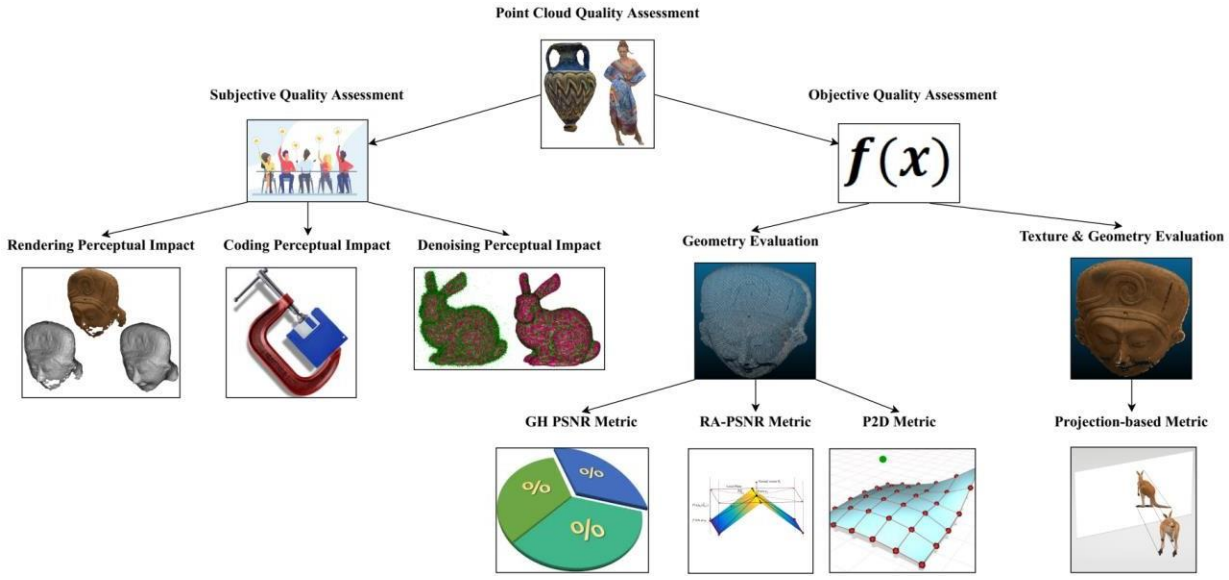



**Figure 1.3:** Structured representation of the novel contributions in this Thesis.

**Part II – Subjective Quality Assessment Studies (Chapters 3, 4, and 5)**

In the context of point cloud subjective quality assessment studies, this Thesis brings the following contributions: i) Subjective assessment methodology for point cloud denoising algorithms; ii) one of the first subjective studies on coding artifacts: iii) performance evaluation of the objective quality metrics available in the literature for different degradations; iv) one of the first subjective studies using the MPEG point cloud coding standards; and v) the only subjective study considering the impact of rendering on point cloud quality. A summary of the novel contributions is described in the following.

**<u>Subjective and Objective Quality Evaluation of 3D Point Cloud Denoising Algorithms</u>**

In this work, the impact of two different types of noise and denoising algorithms on the point cloud geometry quality was subjectively assessed. Then, the performance of available state-of-the art quality metrics for the evaluation of noisy and denoised point clouds was compared. This work offers two main novel contributions:

- **Subjective assessment methodology for point cloud denoising algorithms:** This assessment assumed arbitrary and unstructured point clouds, without an underlying connectivity, what is rather important since point clouds are rendered before being displayed and the presence of noise can lead to annoying artifacts which lead to lower perceived quality. To perform this study, classical image denoising solutions, e.g., using Tikhonov and Total Variation regularization functions, were adapted to point clouds, inspired by the emerging graph-based signal processing field to either remove erroneous points or improve its positioning.
- **Study on the correlation of available objective quality metrics with user subjective scores:** As reliable objective metrics, well correlated with human perception, are urgently needed, this

work performed subjective experiments with the aim to evaluate some selected point cloud objective quality metrics. As far as the authors know, this was the first study of this type made available in the literature and could have a key role on the future development not only of point cloud denoising algorithms but also other types of point cloud processing algorithms. This study considered two types of objective metrics, notably point-to-point and point-to-plane metrics. A Double-Stimulus Impairment Scale (DSIS) protocol has been used for the subjective assessment.

This research work led to the following publication [9]:

- **A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Subjective and objective quality evaluation of 3D point cloud denoising algorithms," *International Conference on Multimedia & Expo Workshops*, Hong Kong, July 2017.

**Subjective and Objective Quality Evaluation of Decoded Point Clouds**

In this work, the impact of two different coding solutions on point cloud geometry quality was subjectively assessed and the performance of well-known objective quality metrics was studied. Two popular point cloud coding solutions were used to code point clouds with different approaches: octree-based coding and simple transform-based coding, notably designed using emerging graph-based signal processing; the color attribute was not coded. The main contributions are:

- **Subjective quality assessment of point clouds with coding artifacts:** A DSIS protocol has been used for subjective assessment. For the first time, point clouds from the MPEG repository were used for a subjective study and two codecs were used to create test materials: an octree-based and a graph-based codec.
- **Objective quality metrics evaluation for point clouds with coding artifacts:** Two types of objective quality metrics for the point cloud geometry were considered, notably point-to-point and point-to-plane metrics. To the best of our knowledge, this was the first study of the correlation performance of objective quality metrics with subjective scores for decoded point cloud data. Therefore, the conclusions of this paper could have an impact on the design of future point cloud codecs.

This research work led to the following publication [10]:

- **A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Subjective and objective quality evaluation of compressed point clouds," *IEEE Workshop on Multimedia Signal Processing*, Luton, UK, October 2017.

**Point Cloud Rendering after Coding: Impacts on Subjective and Objective Quality of Geometry**

The objective of this work was to study the impact of the different artifacts produced by state-of-the-art point cloud geometry codecs for different types of point cloud rendering by performing

subjective experiments and assessing objective quality metrics in several scenarios. This was the first time that the rendering and coding processes, which play a major role on the final perceived quality, were jointly evaluated in this case for static point clouds. In this context, the main contributions are:

- **Point cloud rendering after coding – subjective quality assessment:** Study of the subjective quality impact of multiple rendering-coding combinations for relevant, lossy point cloud geometry coding and rendering solutions. Moreover, the visibility of the distortions associated to each codec under different rendering scenarios was analyzed. This first contribution is critical for the design of suitable point cloud subjective assessment methodologies, where a rendering solution must be chosen.
- **Point cloud rendering after coding – objective metrics assessment:** Evaluation of the subjective correlation performance of available point cloud objective quality metrics for multiple rendering-coding combinations, i.e., for different types of rendering and coding artifacts. This should allow understanding the strengths and weaknesses of available point cloud geometry objective quality metrics as well as their scope of validity, i.e., for which conditions these metrics represent well enough the human perceived quality. This second contribution is critical for the design of more reliable point cloud objective quality metrics, notably for the evaluation of new point cloud geometry coding solutions as well as associated processing techniques.

This research work led to the following publication [11]:

- **A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Point Cloud Rendering after Coding: Impacts on Subjective and Objective Quality," *IEEE Transactions on Multimedia*, (Early access), November 2020.

**Part II – Proposed Objective Quality Assessment Metrics (Chapters 6, 7, 8, and 9)**

Regarding the objective quality assessment metrics, this Thesis proposes four different point cloud objective quality metrics, three of them evaluating only the geometry quality and the last one evaluating texture and geometry together.

**<u>Generalized Hausdorff Distance PSNR-based Quality Metric (GH-PSNR)</u>**

This work proposed a novel point cloud geometry quality assessment metric based on a generalization of the Hausdorff distance. To achieve this goal, the so-called generalized Hausdorff distance for multiple rankings is exploited to identify the best performing quality metric in terms of correlation with the MOS scores obtained from a subjective test campaign. The experimental results showed that the objective quality metric associated to the classical Hausdorff distance leads to low objective-subjective correlation and, thus, fails to accurately evaluate the quality of decoded point clouds for emerging codecs. However, the quality metric associated to the generalized

Hausdorff distance, with an appropriately selected ranking, outperforms the MPEG adopted point cloud geometry quality metrics when decoded point clouds with different types of coding distortions are considered.

This research work led to the following publication [12]:

- **A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "A generalized Hausdorff distance based quality metric for point cloud geometry," *IEEE International Conference on Quality of Multimedia Experience* (QoMEX), Athlone, Ireland, May 2020.

**Resolution-adaptive PSNR-based Quality Metric (RA-PSNR)**

In this work, novel improved PSNR-based point cloud geometry quality metrics are proposed by exploiting the intrinsic point cloud characteristics and the rendering process that must occur before visualization. Point cloud s are rendered on 2D screens to be shown to the subjects. Due to the way point clouds are acquired and pre-processed (before coding), the intrinsic resolution, a measure of distance between points in the 3D space, plays an important role on the final perceived quality, not only to mitigate or highlight coding artifacts but also to measure the intrinsic point cloud quality (i.e., after acquisition). To achieve higher correlation with perceived quality, the proposed objective quality metric considers the point cloud intrinsic resolution as well as the rendering resolution. The rendering resolution corresponds to the resolution of the point cloud after 2D projection for rendering. The main contributions of this work are the following:

- **Resolution-adaptive PSNR:** Proposal and evaluation of several novel point cloud geometry quality PSNR-based metrics that exploit the *intrinsic characteristics* of a point cloud. In this case, intrinsic resolution and precision were considered the most important intrinsic characteristics influencing the final point cloud quality. The main novelty regards the new intrinsic resolution estimators.
- **Rendered-resolution-adaptive PSNR:** Proposal and evaluation of a novel point cloud geometry quality PSNR-based metric that exploits the way point clouds are typically rendered. In this case, the intrinsic resolution is also considered but after rendering. This allows to significantly increase the objective quality metric performance, i.e., to obtain higher correlation with human opinion scores.

This research work led to the following publication [13]:

- **A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Improving PSNR-based quality metrics performance for point cloud geometry," *IEEE International Conference on Image Processing* (ICIP), Abu Dhabi, UAE, October 2020.

**Mahalanobis-based Point-to-Distribution Metric for Point Cloud Geometry Quality Evaluation (P2D)**

To obtain an efficient quality metric, the underlying surface of the reference point cloud should be modelled and the distance of the decoded points to this surface measured. The MPEG D2 point-

to-plane metric models this surface as a plane centered on a point with orientation defined by its normal. However, this is a rather limited approach considering the unstructured nature of the point clouds, especially for points that are not well modelled by this plane. In this work, a novel objective quality metric is proposed which considers a point to distribution correspondence rather than a point-to-point correspondence between two point clouds. In this context, the Mahalanobis distance is adopted to compute the distance between a point and a set of points in the two point clouds under comparison. This metric takes benefit from the fact that several points have more statistical information about the underlying surface than a single point. Thus, the main contribution of this work is a novel point-to-distribution metric considering a new type of correspondence, namely between a point and a distribution of points corresponding to a small point cloud region. The idea behind this novel type of metric is to statistically characterize the point cloud local surface, using the covariance matrix of points within some local region.

This research work led to the following publication [14]:

**A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Mahalanobis Based Point to Distribution Metric for Point Cloud Geometry Quality Evaluation," *IEEE Signal Processing Letters*, vol. 27, pp. 1350-1354, July 2020.

**A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, " A Point-to-Distribution Joint Geometry and Color Metric for Point Cloud Quality Assessment," **Submitted to** *IEEE Workshop on Multimedia Signal Processing,* Tampere, Finland, October 2021.

**Projection-based Point Cloud Quality Assessment Metric**

In the literature, a few works exploited the idea of measuring the point cloud quality by projecting the point cloud into one or more 2D images, i.e., by converting a 3D point cloud into several 2D images. These 2D images can be obtained by performing a projection for different viewpoints, i.e., using different projection centers. Then, powerful 2D quality metrics can be used, to assess the entire point cloud quality by assessing the quality of these projected views. However, the *projection-based metrics* available in the literature were not showing better correlation performance than point-to-point quality metrics, where correspondences are established in the 3D space and errors in the points position and color are accounted.

In this work, a novel projection-based point cloud quality metric is proposed, which addresses the limitations of the state-of-the-art projection-based metrics and achieves high performance in terms of objective-subjective correlation. The main contributions associated to this quality metric are:

- **Geometry alignment:** The idea is to compare the projected reference and degraded color images for two fixed levels of geometry, for both the reference and decoded point cloud geometry. The two scores obtained for both geometry levels are fused together, implying that the proposed approach implicitly considers geometry distortion as well as texture distortion. By comparing images with the same geometry, any misalignment is avoided, which is the main

responsible for the loss of performance in previous state-of-the-art projection-based quality metrics.

- **Padding:** This novel quality metric can handle decoded point clouds with a larger or a smaller number of points with respect to the reference, by performing padding in the 2D domain, this means using interpolated color values instead of a fix background value or even skipping these additional points.

This research work will be submitted to a journal.

**A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Projection-based Point Cloud Quality Assessment Metric," Submitted to *IEEE Transactions on Multimedia*.

**Benchmark of the State-of-the-art Point Cloud Quality Metrics**

This thesis presents an exhaustive survey of all PC quality assessment studies and objective quality metrics available in the literature (Chapter 2) and therefore provide a deeper insight and analysis into the underlying principles and techniques used in the context of this Thesis. Also, it provides an assessment of relevant PC quality metrics literature and its performance, and therefore a good understanding of the overall landscape. The performance of the state-of-the-art metrics are benchmarked in Chapter 9. Point cloud objective quality metrics are also classified in different type of metrics and the performance of almost thirty different quality metrics are evaluated with a relevant dataset.

## 1.4 Thesis Structure

This Thesis reports the main technical contributions developed, notably three subjective quality assessment studies and four new objective quality metrics, each of them outperforming the state-of-the-art point cloud quality metrics in the literature at the time they were developed and published. Figure 1.4 shows the structure of this Thesis.

**Part I - Background**

This part introduces the context and objectives of this Thesis and reviews the relevant background work, notably on point cloud quality assessment; this part includes two chapters.

**Chapter 1** presents the context and motivation for this Thesis, followed by the objectives and main novel technical contributions.

**Chapter 2** reviews the background work for this Thesis by first reviewing the main available point cloud coding solutions and their associated artefacts; after, and more importantly, this chapter reviews the state-of-the-art on point cloud quality assessment.

**Part II - Subjective Quality Assessment and Associated Objective Metrics Assessment**

This part presents the contributions of this Thesis related to subjective quality assessment and includes three chapters.

**Chapter 3** introduces the subjective and objective study performed to evaluate point cloud denoising algorithms and the performance of state-of-the-art objective metrics to assess the quality of denoised point clouds.

**Chapter 4** presents a similar study targeting coding artifacts.

Finally, **Chapter 5** studies the subjective impact of several rendering approaches on the quality of decoded point clouds. The performance of the state-of-the-art objective quality metrics in assessing point clouds coded with the MPEG point cloud coding standards, is also evaluated, and benchmarked in this chapter.

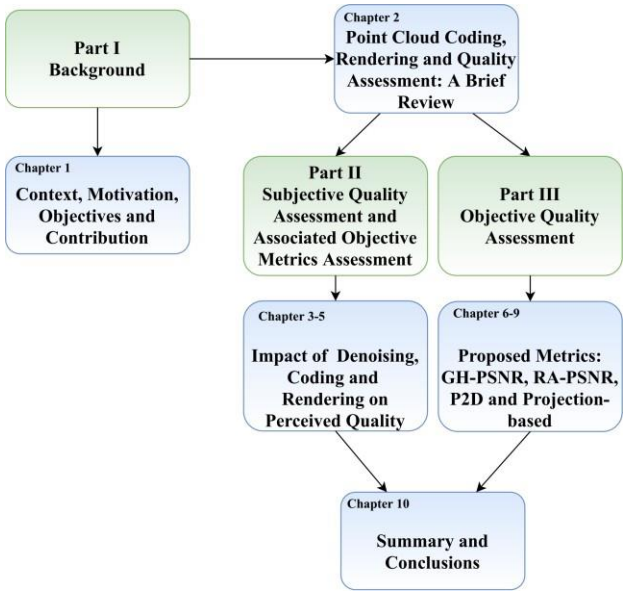



**Figure 1.4:** Thesis structure.

**Part III - Objective Quality Assessment**

This part includes all the novel point cloud objective quality metrics proposed in this Thesis.

**Chapter 6** proposes a metric based on the Generalized Hausdorff distance, which performs well by filtering some of the larger errors.

**Chapter 7** proposes an improvement to the PSNR-based metrics (largely used in the MPEG and JPEG standardization groups) for point cloud geometry, which explicitly considers the point cloud precision and intrinsic resolution.

**Chapter 8** proposes a point-to-distribution quality metric, which considers a new type of correspondence to overcome the weaknesses of state-of-the-art quality metrics to evaluate decoded point clouds with a larger number of points than the reference point cloud.

Finally, **Chapter 9** proposes an image-based metric where the point cloud is projected onto several 2D planes to jointly assess the geometry and color quality using conventional 2D image quality metrics.

Finally, **Chapter 10** closes this Thesis with a summary of the achievements, main conclusions, and some directions for future work.

# Chapter 2

# Point Cloud Coding, Rendering and Quality Assessment: A Brief Review

## 2.1 Introduction

With the recent progress in acquisition sensors and signal processing techniques, a high-resolution 3D model of any object can be acquired in real-time, including animated objects. This visual representation format enables many emerging applications such as immersive communications, cultural heritage, autonomous driving, and virtual and augmented reality. The perceived quality of the obtained point clouds is highly affected by the acquisition method and sensor accuracy. Many applications such as cultural heritage and immersive communication require point clouds with high quality/fidelity comparing to the real-world model. Point clouds are easy to acquire and process; however, they typically contain a large number of points with their associated attributes such as color, reflectance and normal vectors. This large amount of data is expensive to store and transmit and will be a bottleneck for many storage and transmission systems. Naturally, point cloud coding solutions are used to reduce the bit rate while keeping the content quality/fidelity as much as possible. In this context, point cloud quality assessment is essential to design efficient coding solutions and eventually optimize some specific parts of the point cloud coding engine. At the client side, the decoded point clouds have to be rendered before being visualized using a specific type of visualization. Different types of rendering may create different impressions of quality and reality for the point cloud. For decoded point clouds, rendering has different impact on visibility of distortions associated depending on the used coding scheme. Finally, the way users interact with the rendered point clouds is also important for the perception of quality. A user watching a rendered 2D video of a point cloud has a different quality perception from a user who is free to

interact with the point cloud using a head-mounted VR headset. More importantly, in a point cloud communication system, it is rather important to monitor the point cloud quality offered to the users.

The main objective of this chapter is to introduce the main concepts necessary to understand this Thesis and briefly review the state-of-the-art on the key technologies, especially point cloud coding, rendering and quality assessment.

In this chapter, the point cloud processing pipeline is introduced in Section 2.2 to make clear which are the main modules in a typical point cloud application and which factors have a major impact on the perceived point cloud quality. Point cloud coding and point cloud rendering are reviewed in Sections 2.3 and 2.4, as two most important parts of the pipeline with focus in this Thesis. Then, the state-of-the-art on point cloud subjective and objective quality assessment is reviewed in sections 2.5 and 2.6. Finally, in Section 2.7, some final remarks for this chapter will be presented.

## 2.2 Point Cloud Processing Pipeline

A typical point cloud processing pipeline is shown in Figure 2.1, including acquisition, representation, encoding and decoding, rendering and visualization modules. It is important to highlight that every single module in the pipeline may affect the point cloud quality in different ways. In Figure 2.1, for better illustration, modules include some images inside them.

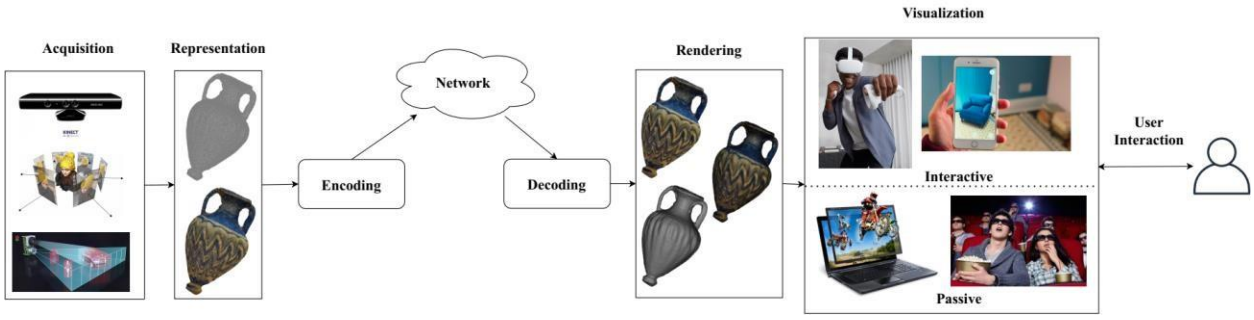



**Figure 2.1:** Point cloud processing pipeline.

The various modules in the point cloud processing pipeline are as follows:

- **Acquisition**: The 3D representation of the objects in real world is captured with appropriate sensors, such as depth sensors, laser scanners (LIDAR), and arrays of cameras which can be arranged differently. This means that based on acquisition setup, point cloud data can be acquired directly as a set of 3D coordinates, using laser scanners or indirectly as a set of 2D images or depth maps, eventually with some associated texture information. The perceived quality of a point cloud strongly depends on the acquisition accuracy and reliability of the sensors, which can introduce errors and noise in the 3D coordinates.
- **Representation**: While 3D data is acquired with a specific model, depending on the available sensors, it may happen that this is not the best representation model for the target application and some representation model conversion needs to be applied before coding. The most

common example is to perform the acquisition with array of 2D cameras images, basically a light field, and convert this data into a point cloud using appropriate signal processing techniques.

- **Encoding**: The encoding module is essential to deal with large point clouds, notably with up to billions of point coordinates and their associated attributes. The way coding is performed has a strong impact on the perceived quality of point clouds. Moreover, it may lead to artefacts that are very different than the artefacts typical of 2D images. Due to its importance, the next section will briefly review the point cloud coding state-of-the-art, especially the recent MPEG standards on (lossy) point cloud coding.
- **Decoding**: The bitstream at the client side is processed to obtain the decoded point clouds. The decoded point cloud is typically different from the reference point cloud in terms of the precise position of points as well as errors in the color components which are typically introduced in the quantization process. Moreover, the number of decoded points, their precision (or bit-depth) may not be the same.
- **Rendering**: Based on the type of visualization intended and the application requirements, point clouds may be rendered in a different way. There are two main approaches for rendering: points may be rendered directly with a primitive, i.e., using some elementary structure, or some connectivity may be associated to the points to create a surface (e.g., a polygonal mesh) with the rendering performed with these surfaces. As expected, the rendering has a critical impact on the perceived quality of the decoded point clouds. Due to its importance, Section 2.4 briefly reviews point cloud rendering approaches, especially the most common ones, and discuss their advantages and disadvantages.
- **Visualization**: Point clouds are rich 3D visual representations and thus can be visualized with different devices with (interactive) or without interaction with the users (passive). The simplest approach is to render point clouds on a 2D screen from some specific viewpoint (that can be changed by the user), but other devices are also possible, such as a stereoscopic display (two views) with passive glasses or auto-stereoscopic displays (tens or thousands of views). More importantly, virtual and augmented reality displays can also be used to display point clouds, eventually with other representations (e.g., meshes) or embedded in the physical word. In this last case, user interaction is more natural since it follows the head and/or body motion.

From all processes applied to a point cloud, *point cloud coding* and *rendering* have the most significant impact on the perceived quality as it will be explained in more detail next.

## 2.3 Point Cloud Coding

Point clouds may have billions of points and their associated attributes; for this reason, coding is critical before storage and transmission. There are several coding approaches for point cloud coding in the literature. Considering the scope of this Thesis, the most relevant and representative point cloud coding solutions available in the literature, later used for subjective and objective

quality assessment, are reviewed in this section. Naturally, the MPEG G-PCC [3] and V-PCC [4] codecs, which were recently standardized, are the most relevant for this work as they are the most recent and efficient point cloud coding solutions available. These codecs are part of the MPEG-I set of standards, which aim to design key technologies for immersive media. The so-called Point Cloud Library (PCL) codec is the first point cloud coding scheme which became popular and adopts an octree-based coding approach.

Since the PCL [15], MPEG G-PCC, and V-PCC codecs are often used in many parts of this Thesis, they will be reviewed here. These codecs represent the three most relevant approaches to structure point cloud data for coding purposes, namely based on tree, surface, and patch, respectively. A tree is a data structure used to organize the points in some hierarchical way, e.g., octrees, kd-trees; a surface is a data structure where the points are represented with a parametrized surface model (e.g., as a set of triangles); finally, a patch clusters points into groups with some size, which is suitable for 3D to 2D projections. In this Thesis, only coding of geometry and color of static point clouds is considered [16]. Naturally, these point cloud codecs produce different types of geometry artifacts, such as loss of geometric detail, geometric deformations, holes creation and other geometric distortions, e.g., when curved surfaces are represented by a small number of planes as well as color artifacts such as blocking artefacts, etc.

### 2.3.1 Point Cloud Coding with Tree Structures

The point cloud coding solution selected for this class is the popular point cloud codec, publicly available in the Point Cloud Library [15], a large scale, open project for 2D/3D image and point cloud processing. To facilitate the coding of geometry data, this codec represents the point cloud 3D coordinates and its attributes with an octree structure [17]. The PCL point cloud codec is often used as performance benchmark since it can handle unorganized point clouds of arbitrary size/density which may be acquired with different types of sensors (e.g., LIDAR or stereo based cameras) and has low encoding and decoding complexity. Figure 2.2 illustrates the block diagram of the PCL codec.

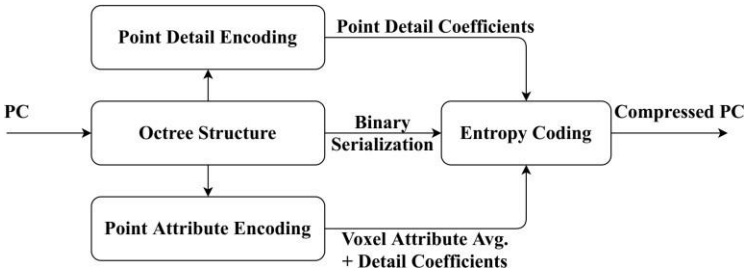



**Figure 2.2:** PCL Static point cloud encoding architecture **[17]**.

The key modules of the PCL codec are:

- **Octree Structure:** In PCL, each octree node corresponds to a voxel in 3D space. The root node corresponds to a voxel that contains all points of the point cloud, the so-called *point cloud bounding box.* Then, starting from the root node, each voxel is divided iteratively into 8 voxels with the same size; however, a node is not divided if the corresponding voxel is not occupied. The occupancy of a node is represented with a single byte that signals the occupied child nodes. By traversing the octree in breadth-first order, a stream of occupancy bytes is created, thus allowing an efficient representation of the point cloud geometry. This process is called serialization. Figure 2.3 shows an example of the relationship between an octree and its byte stream, created via serialization. The decoded quality is determined by the octree depth, which indirectly specifies the minimum voxel size; when the octree does not have enough depth to represent the points with their precision, it corresponds to a pruned octree. When the point cloud is decoded, all the points inside an occupied voxel are represented with just one point at the voxel center.

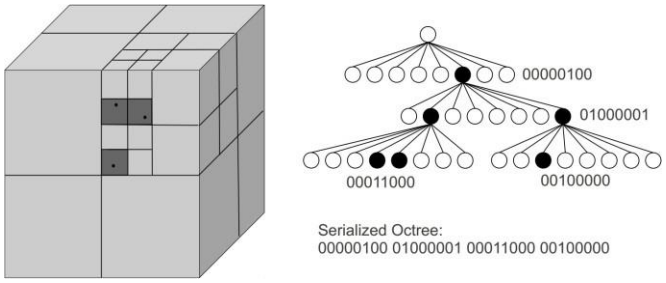



**Figure 2.3:** Overview of an octree structure and its serialization.

- **Point Detail Coding:** This process aims to improve the point coordinates accuracy without further increasing the octree depth. In this step, all the points inside a voxel at a specific octree depth are differentially encoded with respect to the center of the voxel. If point detail coding is disabled, all points inside a voxel are represented by the voxel center point. This means that the number of decoded points is reduced compared to the reference.
- **Point Attribute Coding:** Once an octree structure is generated, during the pruning, the color associated to the voxel center point is the average color of all the points in that voxel. In case point detail coding is enabled, the same approach used for geometry is also applied to the point cloud attributes and detail coefficients are calculated, which correspond to the difference between the voxel points average color and the reference color value of each point.
- **Entropy Coding:** The statistics of the occupancy bytes are exploited by an entropy encoder that considers the specific (non-uniform) symbol frequencies to reduce the final bitrate. The entropy coder for this codec is a range coder [18] which is variant of an arithmetic coder.

### 2.3.2 Point Cloud Coding with Surface Models

The point cloud coding solution selected to represent this coding approach is the MPEG G-PCC codec, which is capable of lossy and lossless coding of large point clouds, with spatial random access, view dependent processing, packetization, and scalability [19]. As the PCL octree-based codec, the G-PCC codec is also based on an octree decomposition to code the point cloud geometry but extends this coding approach with a parameterized surface model. As for PCL, a pruned octree is used but the geometry of the points at each leaf voxel is not represented by the voxel center; instead, a set of triangles is used to represent a surface formed by these points. Figure 2.4 shows the architecture of MPEG G-PCC encoder.

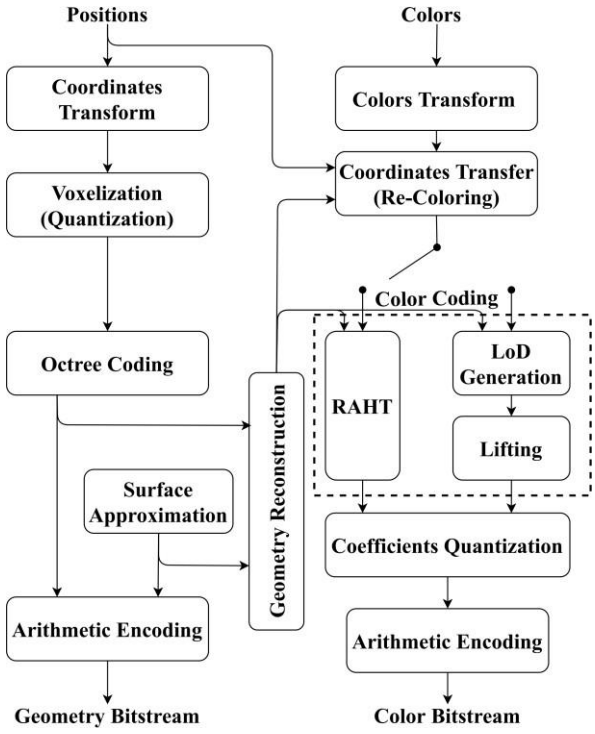



**Figure 2.4:** MPEG G-PCC encoder architecture.

The key modules of the G-PCC codec are:

- **Coordinate Transform and Voxelization:** The points coordinates are first transformed to lie in the cube $[0, 2^d - 1]^3$, where $d$ corresponds to the octree (full) depth parameter (defined *a priori*). Then, they are truncated so that each point is represented by the center of the voxel which is a non-negative integer (voxelization process).
- **Octree Coding:** An octree structure is created for the point cloud data and then pruned from the root down to some specific octree level $l$, which must be smaller or equal than the octree depth; for lossless coding, the octree level must be equal to the octree depth.

- **Surface Approximation:** If $l$ is smaller than depth, a polygonal representation is used to represent the points, which is known as *TriSoup*, an amalgam for Triangle Soup. This means that the limited depth octree is complemented with additional geometry information within groups of voxels, called *blocks*; this additional geometry is represented by vertices, corresponding to the intersections of the surface with some edges of the block (in this case at most 12 vertices). This set of vertices is sufficient to reconstruct a surface, corresponding to a non-planar polygon passing through all the vertices. A bit vector determines which edges contain a vertex (i.e., intersection with the block edges) and which do not; then, for each edge containing a vertex, the position of the vertex along the edge is uniformly scalar quantized. The bit vector and the quantized positions are transmitted to the decoder side.
- **Geometry Reconstruction:** Since color attributes must be coded and transmitted only for the decoded points, the geometry needs to be decoded; this needs to be performed before the coding of the color or other attributes and the same operation occurs at the decoder. In this module, the octree is decoded, and then a surface is reconstructed from the position of vertices and bit vector by surface approximation. A decoded point cloud is created with a similar resolution (number of points per volume) to the reference point cloud or more by sampling the reconstructed surface.
- **Color Transform:** Color of points may have to be transformed from RGB to $YC_bC_r$ color space.
- **Color Transfer (or Re-Coloring):** For lossy geometry coding, the reconstructed points are recolored using the color information available for the reference point cloud. In G-PCC, each decoded point is assigned with the color of the closest reference point in terms of Euclidean distance.
- **Color Coding:** There are three color coding modes in G-PCC, notably Region Adaptive Hierarchical Transform (RAHT) [20], Predicting Transform, and Lifting Transform [19]. These coding modes are explained next:
    - **RAHT:** Color data is spatially transformed using RAHT which is an adaptive form of the Haar wavelet transform, performed with respect to a hierarchy defined by the so-called *Morton codes* (i.e., indices) of the voxels. *RAHT* is recursively applied from the target depth (level) up to the octree root, to blocks of voxels grouped two by two at each level and along each of the three coordinates successively.
    - **Predicting Transform (LoD Generation):** The *Predicting Transform* is an interpolation-based hierarchical nearest-neighbor prediction scheme for attribute coding [19], implying that already encoded colors are used for prediction. It relies on a Level of Detail (LoD) representation that distributes the input points into sets of refinements levels (R) using a deterministic distance criterion. The attributes of each point are encoded by using a prediction determined by the LoD order. The maximum number of

prediction candidates can be specified, and the number of nearest neighbors is determined by the encoder for each point.

- **Lifting Transform:** The *Lifting Transform* [19] is built on top of the *Predicting Transform*. It introduces an update (Lifting) operator and an adaptive quantization strategy. In the LoD prediction scheme, each point is associated with an influence weight. The prediction residuals are used to update the color/attribute values at each stage of the prediction and guide the quantization process.

- **Coefficients Quantization:** Transformed color coefficients are quantized by a uniform quantizer.
- **Arithmetic Coding:** Geometry symbols and quantized transform color coefficients are encoded using an arithmetic entropy coder, thus building the corresponding bitstream.

At the decoder side, after arithmetic decoding, the octree is reconstructed, and surface reconstruction is performed using the decoded vertex information and bit vector. Then texture information is inverse transformed to obtain the color value for each decoded point.

### 2.3.3 Point Cloud Coding with Patch-based Projection

The point cloud coding solution selected for the patch-based coding approach is the MPEG V-PCC codec, which targets dynamic point cloud coding and performs a 3D to 2D mapping of both the geometry and color components [21]. Thus, depth and texture images are created which may be coded with any video codec, notably a High Efficiency Video Coding (HEVC) standard-compliant codec [22]. Figure 2.5 shows the architecture of the MPEG V-PCC encoder.

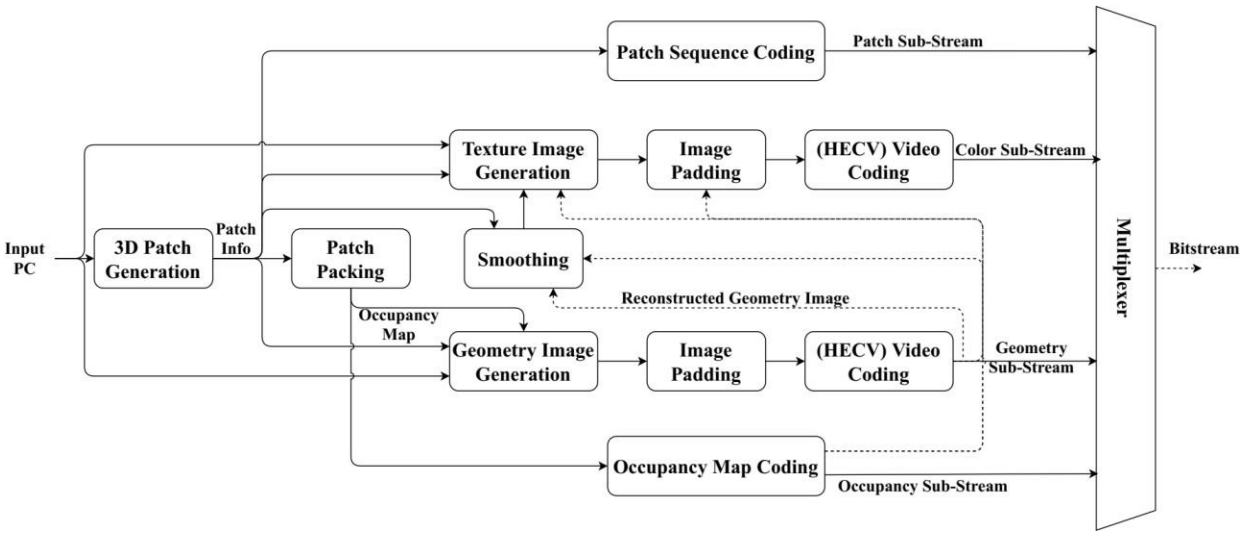



**Figure 2.5:** MPEG V-PCC encoder architecture.

The key modules of the V-PCC codec are:

- **3D Patch Generation:** In the first step, the point cloud is decomposed into several patches with smooth boundaries, while minimizing the reconstruction error. Point cloud points are clustered according to the relation between their normal vectors and the normal directions of six predefined oriented planes (forming a 3D bounding box). Typically, this step includes several operations such as normal estimation, clustering according to the normal vectors and six pre-defined planes (*xy*, *-xy*, *yz*, *-yz*, *xz*, *-xz*) and a connected components algorithm.
- **Patch Packing:** Patches are extracted from the clusters created in the previous module and mapped onto a 2D grid using a packing process which attempts to minimize the unused space. At the end, each N×N (e.g., 16×16) block in the grid is associated with a unique patch.
- **Geometry Image Generation:** After the packing where geometry (depth) is mapped from 3D to 2D. The 3D point positions are represented in the gray-level (i.e., monochromatic) geometry images. Considering projection on the *xy* plane, while the $x$ and $y$ coordinates correspond to the position of the pixels within the frame, the remaining $z$ coordinate corresponds to the pixel intensity value, similarly to a depth map. To handle multiple points being projected on to the same pixel, each patch is projected onto two geometry images, referred to as the *near layer (D0)* and *far layer (D1)*, containing points with the lowest and largest depth ($z$) values, respectively. There are several approaches to code the *D0* and *D1* layers: i) *D1* is coded differentially with respect to *D0*, thus describing the surface thickness; ii) using a single layer where *D0* and *D1* are interleaved across the rows or columns of the image; iii) *D0* and *D1* are sub-sampled and spatially interleaved; and iv) *D1* is dropped and interpolated at decoder side.
- **Texture Image Generation:** Since the decoded geometry may have a different number of points, texture image generation requires the decoded geometry to compute the colors associated to the decoded points. A recoloring process transfers the color information from the reference point cloud to the decoded geometry point cloud. The pixels in the texture images take the color values of the mapped points in a similar process to the previous step, thus texture images are created.
- **Smoothing:** Patch boundaries and reconstructed geometry images are smoothened using some filters. For geometry, the idea is to avoid discontinuities that may appear at the patch boundaries. In this case, boundary points are moved to the centroid of their nearest neighbors. For color, adaptive filters based on the median, averaging, bilateral filter, order statistic filter can be used. The smoothing process is signaled to the decoder side.
- **Image Padding:** The empty spaces between the patches packed inside the generated images are filled using a padding process to obtain a smoother image (easier and cheaper to code). This will enable an efficient image or video compression and minimize block artifacts. For geometry images, an adaptive padding strategy is performed at $T \times T$ block level: i) for a block with all occupied positions, nothing is done; ii) for a block with all non-occupied positions, the pixels in the block are filled with the values of the last block row/column; and iii) for a partially filled block, empty pixels are iteratively filled with average values of neighboring pixels. For

texture images, an optimization procedure finds the values of empty pixels such that the obtained padded image is as smooth as possible.

- **(HEVC) Video Coding:** These above created maps are passed to some video encoder, e.g., an HEVC encoder, which exploits the spatial and temporal correlations in a very efficient way. This step is the core of encoding process for the V-PCC encoder. For this, the 10-bit HEVC profiles are preferred since they can improve the accuracy and compression efficiency.
- **Patch Sequence Coding:** To reconstruct the 3D data from the 2D maps at the decoder side, for each patch, the index of the unique patch and the index of the projection plane associated to its normal vector direction is encoded, as well as its 2D bounding box and 3D location. 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.
- **Occupancy Map Coding:** An occupancy map is used to determine whether some $B_0 \times B_0$ block is occupied or not; this map is coded to determine which 3D points should be decoded/filled. This binary map can be lossy or lossless coded and $B_0$ is a user-defined parameter which can control the trade-off between precision and rate. For lossless coding, this information is coded directly using binary arithmetic coding. For lossy coding, the occupancy map is coded as a video frame that may have a smaller spatial resolution than the geometry and texture frames.

At the decoder side, four different coded streams are demultiplexed and decoded, namely the patch information, the occupancy map and the geometry images are decoded first and then the texture information is decoded to obtain the color attribute for each decoded point.

## 2.4 Point Cloud Rendering

Point cloud rendering is the process of producing a visual representation that can be consumed by the users using an available display, e.g., conventional 2D, stereo, auto-stereoscopic, head-mounted displays, etc. [23]. Since it effectively selects the information to be seen, the rendering process has a significant impact on the quality perceived by the user. In this section, the rendering solutions selected for the experiments reported in the Thesis are briefly described.

Regarding point cloud rendering, there are two main approaches: the first, so-called *point-based*, directly uses the point cloud data while the second, so-called *mesh-based*, converts the point cloud data into another representation format, very commonly a surface such as a polygonal mesh. The decision on the rendering approach to adopt mostly depends on the application requirements which may be very different.

The point cloud conversion to another more rendering friendly representation format may bring some loss of information and, in some cases, it may not even be possible due to the complexity of the visual scene in terms of geometry or the low point cloud density. By directly rendering point clouds, massive amounts of points can be visualized. Rather often, these point clouds do not fit

into the available memory and require special algorithms to stream, process and render only a small subset of the entire point cloud data. This is easier to perform with a point-based model due to the lower complexity associated to the rendering process in comparison to a polygonal mesh representation where surface reconstruction and interpolation are usually needed. In point-based rendering, the so-called *rendering primitives*, the simplest (atomic) elements, are combined to create the 3D impression of a surface in the final displayed content. 2D rendering primitives may be circles (splats), squares, quads, etc. and 3D rendering primitives are cubes, spheres, etc. Typically, large scale models for geographical information systems are usually rendered with a point-based representation while mesh-based representation may be preferred for immersive communications.

Independently of the rendering approach, a geometry shader with some primitives is employed to construct the final image shown to the user. In this context, the geometry shader is responsible for the creation of appropriate levels of light, darkness, and color within a rendered image/view [24]. For point clouds only with geometry, the shading is commonly performed with a single color; otherwise, color attributes are used for each point or vertex.

### 2.4.1 Point-based Rendering without Color Attributes

Point-based rendering algorithms use a set of discrete points that may be irregularly distributed, simple rendering primitives and 3D/image space interpolation procedures to obtain a rendered view, i.e., a 2D image. The main advantage of this rendering approach, labelled here as *RPoint*, is that it can achieve high levels of realism and is adequate for complex objects, such as trees, feathers, smoke, water, etc. In addition, point-based rendering simplifies the rendering process and typically requires less memory and computational power due to the lack of connectivity information.

In this approach, depending on the visualization device and type of rendering, simple and fast to render 3D or 2D primitives are selected, such as circles (splats), squares, quads for 2D rendering and spheres, and cubes for 3D rendering. Based on the point cloud density and distance to the virtual camera (zoom level), the size of the primitives can be manually or automatically adjusted to create the impression of a surface; in the automatic case, connectivity information between points is usually computed to determine the appropriate primitive size [25]. The definition of an appropriate size for the primitive is rather important to reduce the appearance of empty spaces (holes) between points (size too small) or aliasing artifacts (size too large).

Regarding shading, color attributes are not used in this specific rendering approach; this may be useful to assess the impact of geometry distortions without the influence of any additional component if it exists. The human visual system can easily and accurately derive the 3D orientation of surfaces by using the variations on the image intensity alone [26].

### 2.4.2 Point-based Rendering with Color Attributes

The second rendering approach is still point-based but also uses the available color attributes and thus, for this reason, it will be labelled as *RColor* in the following. In *RColor*, the *RPoint* rendering method above is again applied but the point color attributes are used. This means that the surface is still represented with points and displayed with the same primitives but now rendered with the color data obtained during the point cloud acquisition process. While the color attributes correspond to the real color of the objects, they are still influenced by the specific light conditions that have occurred during their acquisition. For this reason, no shading applied to points. However, in the final rendered image, some colors may be interpolated, e.g., between points, to avoid aliasing. Moreover, since the captured color also conveys the object depth, it may mask some geometry distortions on the geometry/surface. On the contrary, distortions may be more visible at object boundaries, which are essential to give the user the shape perception of the objects in the visual scene.

### 2.4.3 Mesh-based Rendering without Color Attributes

The first step in the mesh-based rendering approach, hereafter labelled as *RMesh*, is to create polygonal meshes with a surface reconstruction algorithm, such as the *Poisson Surface Reconstruction* [27]. This means that rendering is performed with a set of vertices along with their connectivity to obtain a closed surface, very precisely defined.

The advantage of this rendering approach is that, independently of the distance to the object (or scene) or the point cloud density, a seamless surface is obtained; this may not occur with point-based rendering since the quality depends on the number of points representing the surface and the distance between the viewer and the object. The key disadvantage of this rendering approach is that it requires surface reconstruction, which usually removes high frequency geometry details [28], i.e., smoother surfaces are obtained. It is important to note that surface reconstruction for complex surfaces is not always straightforward, i.e., it may not always be successful and can even require some user intervention. After surface reconstruction, the polygonal mesh needs to be rendered, usually with some shading algorithm [29] [30]. There are several mesh rendering techniques performing shading, reflection, refraction, and indirect illumination, and able to improve (when properly applied) the quality of the rendered data.

## 2.5 Point Cloud Subjective Quality Assessment

In this section, the most relevant subjective point cloud quality assessment concepts and related works will be reviewed.

### 2.5.1 Background and Key Concepts

In subjective point cloud quality assessment, the degraded point cloud quality is assessed with some well-defined procedures and opinion scores are collected from some observers; these opinion scores may also be used as reference to validate objective point cloud quality metrics. The design

of a subjective quality assessment test requires some basic components that must be selected depending on the application and aim of the test:

- **Test environment:** Subjective tests can be performed in almost any environment. However, due to possible influence from outside conditions, it is typically advised to perform tests in a neutral environment, such as a dedicated laboratory room. Such a room may be sound-proofed, with walls painted in neutral grey, and using properly calibrated light sources. Several well-known recommendations specify these environment conditions [31] [32].
- **Visualization device:** The type of visualization device and its configuration should also be decided depending on the application and purpose of the test. For point clouds, different types of displays can be used, such as augmented or virtual reality displays, stereoscopic or auto-stereoscopic displays and still most commonly classical 2D displays.
- **Test materials:** The test material is the content selected for the subjective experiment. Typically, reference point clouds with different characteristics are evaluated under different degradations, e. g. different coding rates, different noise levels, etc.
- **Test Methodology:** The methodology used to perform the subjective experiment, this means how the stimuli are shown to the observers and how the opinion scores are collected, is very important and needs to be carefully selected depending on the application and purpose of the test. There are several test methodologies, many already defined in suitable standards [31] [32] [33]. The three most popular test methodologies for point cloud quality assessment are:
    1. **Double Stimulus Impairment Scale (DSIS):** Reference and degraded point clouds are shown side-by-side or consecutively in time to the users who must score the level of impairment for the degraded point cloud comparing to the reference point cloud, e.g., in five levels: "imperceptible", "perceptible but not annoying", "slightly annoying", "annoying", and "very annoying". Naturally, these point clouds are rendered with some technique and shown on a specific device. This type of methodology is often used to evaluate visual degradations due to compression, where the reference (or reference) stimuli are available, and the fidelity is the main aspect to be accounted and assessed.
    2. **Absolute Category Rating (ACR):** The ACR method, aka single stimulus method, displays one point cloud at a time and the observers assess the quality without a reference. Each of the point clouds is scored individually, e.g., with five levels: "excellent", "good", "fair", "poor", and "bad". However, the ACR method may be inappropriate since, without a reference, the observers may use their own internal reference and the scores may have some bias. This disadvantage can be mitigated by using the so-called ACR with hidden reference (ACR-HR) method where the reference point clouds are also assessed together with the degraded point clouds.
    3. **Pairwise Comparison (PWC):** Point clouds are shown side-by-side and observers must select the one with the highest quality (preferred), which is a rather simple and fast task.

Unlike the previous test methods/protocols, a psychometric scaling operation is required to convert the results to numerical subjective quality scores. Nowadays, this methodology is regarded as one of the most accurate subjective assessment methods.

- **Test Subjects:** Test subjects must assess the quality and are also called observers. A minimum number of subjects is needed for a subjective assessment test to be statistically meaningful. While a larger number of subjects increases the reliability of the collected experimental data, notably by reducing the standard deviation of the averaged scores, it is usually difficult and even costly to attract participants to the subjective tests. The minimum number of subjects that are required for a subjective video quality study to be statistically meaningful is not very strictly defined. According to ITU-T [31], a minimum of 15 subjects should be used. However, fewer than 15 subjects are also possible for studies with very limited scope. Ideally, subjects should be non-experts, but experts can also be invited if their number is not too high. In any case, the level of expertise of the subjects should be reported.
- **Data Analysis:** After collecting the subjective scores, outlier subjects must be rejected according to the procedures defined in Recommendation ITU-R BT 500.13 [31], notably: i) first, it is determined whether the distribution of scores for a test sequence is according to a normal distribution or not; if the kurtosis (fitness measure) of all subjects scores for each test stimulus/sequence is between 2 and 4, then the distribution can be considered as normal; ii) to reject a subject, a confidence interval is defined using the following procedure: if scores are distributed normally, for each score larger than two standard deviations (higher limit) above the mean score for that test stimulus/sequence, a counter $P_i$ is incremented. Also, for each score smaller than two standard deviations (lower limit) below the mean score, a counter $Q_i$ is incremented. In case of non-normal distributions, the upper and lower limits are $\sqrt{20}\sigma$ from the mean to increment $P_i$ and $Q_i$. For a subject to be rejected, it is necessary to fulfill the following two conditions:

$$\frac{P_i}{N_m} > 0.05 \quad \text{and} \quad \left|\frac{P_i - Q_i}{P_i + Q_i}\right| > 0.5 \tag{2.1}$$

where $N_m$ is the number of stimuli.

After outlier removal, the average of all scores across the subjects are computed for each point cloud stimulus, thus obtaining the overall Mean Opinion Scores (MOS). The MOS for each point cloud stimulus/sequence $c$ will be:

$$MOS_c = \frac{\sum_{i=0}^{N_s} SQS(i, c)}{N_s} \tag{2.2}$$

where $N_s$ is number of subjects in the test and $SQS(i, c)$ is the subjective quality score given by subjects $i$ to the stimulus/sequence $c$; a stimulus typically corresponds to specific point, coded at a specific rate, with a specific codec.

### 2.5.2 Point Cloud Subjective Quality Assessment: State-of-the-Art Review

This section will review the major advances in point cloud subjective quality assessment in recent years.

In [34], Zhang *et al.* have designed a subjective test for point clouds rendered with *RColor* for different levels of degradation, both for geometry and color. The quality degradations have been introduced by down sampling the geometry and adding (synthetic) uniform noise to both color and geometry. The main conclusion was that human perception is more tolerant to color noise compared to geometry noise in point clouds.

In [35], Mekuria *et al.* conduct a subjective assessment experiment using a point cloud codec based on geometry octree pruning and JPEG based attributes coding. The subjective evaluation was performed in a mixed reality system, combining coded point cloud data (acquired) rendered with *RColor* with shading and computer graphics generated 3D content rendered as *RMesh* with color. In the subjective test, the users could interact with the content by navigating a visual scene with an avatar. The system performance was globally assessed with a questionnaire addressing eight different quality aspects, among them realism, impressiveness, and color quality. Two objective quality metrics (mean squared error based) were introduced to assess both the geometry and color qualities. However, the correlation between objective metrics and subjective results was not assessed.

In [36], the performance of the codec presented in [35] is subjectively evaluated again. Both static and dynamic point clouds are evaluated under several coding configurations such as lossy, lossless, all intra, etc. at different rates for dynamic and static point clouds. A passive subjective assessment approach was adopted, and video sequences of the point cloud models rendered with *RColor* were generated from predefined viewpoints. The point clouds were rendered using cubes of fixed size as primitive.

In [5], Alexiou *et al.* perform a subjective quality assessment study of the point cloud geometry for two types of degradation, notably octree pruning and Gaussian noise, considering different quality levels and artifacts. An augmented reality (AR) headset was used to visualize simple point cloud objects without color (*RPoint* without shading) from different perspectives (user could move around the object). It was concluded that the objective quality metrics could perform well for Gaussian noise but underperformed for PCL-like compression artifacts.

In [6] and [37], Alexiou *et al.* perform a subjective test with the same data as in [5] and the same distortion types but the content was visualized on a 30-inch 2D display. In both cases, *RPoint* without shading rendering was used, and user interaction was allowed; a simple rendering method with unit size points was selected. In [37], the impact of adopting two different subjective assessment methodologies, Absolute Category Rating (ACR) and Double Stimulus Impairment Scale (DSIS), was studied by comparing the results obtained. The DSIS protocol was found to be more consistent and with lower confidence intervals and, thus, it was used later in [6]. In [6]it was shown that state-of-the-art objective quality metrics perform well in the presence of Gaussian

noise. However, the performance for predicting the quality of octree-based codecs is content dependent and, therefore, metrics are not performing well for this type of distortion.

In [8], Alexiou *et al.* perform a subjective study to evaluate the point cloud geometry quality using an octree pruning-based codec. Before rendering, a Poisson surface reconstruction algorithm was used to obtain a mesh from the decoded point clouds without color (*RMesh*). In this case, no interaction was allowed with the content and the subjective experiment also followed the DSIS protocol. It was found that most point cloud objective quality metrics have a low correlation with the subjective scores and the 3D surface reconstruction algorithm plays a crucial role on the quality scores obtained.

In [38], Alexiou *et al.* perform two subjective tests to study the impact of visualization on the subjective quality assessment of point clouds. The first test used a 30-inch 2D display and the second an AR headset. As before, geometry artifacts associated to octree coding and Gaussian noise were studied and point clouds are rendered with *RPoint* without shading. The assessment protocol was DSIS and user interaction was not allowed. In any case, the correlation between the subjective scores obtained with different visualization devices was rather high, notably statistically equivalent for the two experiments with Gaussian noise.

In [39], Dumic *et al.* investigate the impact of visualization on the point cloud perceived quality. Several point clouds from the JPEG repository [40] were used and two computer generated geometrical volumes: a cube and a torus. Some point clouds include color and some others only geometry. Reference point clouds rendered as *RPoint*, *RColor* and *RMesh* with and without color are used to evaluate the quality in a 2D display versus a stereoscopic 3D display. They follow a passive approach and use an ACR methodology to conduct the subjective tests in three different laboratories. They show that there is no preference for viewing point clouds in 3D displays over 2D displays.

In [41], Alexiou *et al.* conduct a subjective quality evaluation experiment to assess the quality of decoded point clouds rendered as *RMesh* in several types of 3D displays, from passive stereoscopic to auto-stereoscopic displays. Geometry degradations in the form of octree pruning were evaluated in the absence of color. The results obtained with 3D displays have a strong correlation with the results obtained with 2D displays for the same content. However, it was also found that the rendering method may play a significant role in this evaluation.

Moreover, Alexiou *et al*. have benchmarked objective quality metrics for point cloud data represented by octree pruning and corrupted with Gaussian noise [42]. Both DSIS and ACR assessment protocols were used in separate sessions with point clouds rendered with *RPoint* without shading. It was found that the correlation between subjective and objective quality scores was low for distance-based objective metrics, for octree-based compression artifacts, but better correlation performance could be achieved with metrics considering the normal at each point.

In [43], Christaki *et al.* perform a subjective study for simple point clouds, that were converted to meshes and coded with suitable open-source mesh codecs. While some of the test point clouds are

common with the point clouds often used in previous subjective quality evaluation studies (e.g., Bunny) others were obtained with a platform designed for 3D human capture (with multiple Kinect devices). In [43], a variant of the pairwise subjective assessment protocol was used for evaluation with three stimuli rendered with *RMesh* and presented simultaneously. Overall, three mesh codecs were considered, and content was displayed with a VR application on a head-mounted display. The key conclusions are that usual 3D mesh quality metrics have a low correlation performance in this scenario and the 3D mesh surface reconstruction method plays an important role.

Torlig *et al.* [44] evaluate the visual quality of voxelized colored point clouds rendered with *RColor* in subjective experiments that were performed in two different laboratories. Point cloud voxelization is performed in real-time, and orthographic projections of both the reference and the distorted models are shown side-by-side to the subjects in a platform which allows user interaction. The projected images shown to the user are then used by conventional 2D quality metrics, which allows to measure the perceptual quality of the point clouds in an objective way. The subjective study was only used to assess the performance of the point cloud objective quality metrics.

In [45], Alexiou and Ebrahimi use degraded models as in [44] and the point clouds are rendered with an *RPoint* approach using cubes as primitive geometric shapes. The point clouds are evaluated in an interactive renderer, where the users' behavior is also recorded. The logged interactivity information is further analyzed and used to identify important viewing directions of the point clouds under assessment. The performance of projection-based quality metrics in [35] was improved by exploiting this interactivity data, by assigning weights to the images projected from different views based on the time that users had spent on each one. While one view is enough to achieve high performance, interactivity information can be used to further improve the performance of a projection-based metric.

In [46], Dumic *et al*. review the state-of-the-art on point cloud subjective quality assessment and point cloud objective quality metrics.

Most of the works above mentioned, study the subjective quality impact of point cloud degradations associated to octree pruning and Gaussian noise. More recently, in [47], Alexiou *et al.* perform a subjective evaluation of MPEG point cloud codecs (as well as the MPEG adopted objective metrics) considering both decoded geometry and color. They used three subjective test sessions all using *RColor* rendering; in the first, MPEG common test conditions were used to encode some test material and their quality was evaluated using a DSIS methodology. In this experiment, MOS scores are obtained by data processing which allows to evaluate the performance of the MPEG point cloud codecs and benchmark objective quality metrics. In the second session, as PWC method was used to determine the user preferences among different types of geometry artifacts associated to G-PCC coding configurations, namely between *TriSoup* and *Octree*. Finally, in the last session, a similar PWC test is done to assess which bit allocation is most efficient and visually pleasant between geometry and texture for some target bitrate.

In [48], Zerman *et al.* describe two subjective tests performed with two dynamic point clouds encoded with the V-PCC codec, one using a DSIS methodology to assess large differences in quality and another a PWC methodology to assess smaller quality differences. Point clouds in both sessions are rendered with *RColor*. Four different subsampled versions of the reference point clouds have been decoded with V-PCC and compared together. Experimental results show that the final quality is independent of the number of input points for their dataset.

Zerman *et al.* in [49] have also conducted a subjective study to compare textured mesh and point cloud under different coding methods: G-PCC [3], V-PCC [4] for point clouds and Draco [50] for meshes. They built a dataset with eight dynamic point clouds from 8i [51] and point cloud and mesh data captured by the authors. These point clouds are coded with three codecs at six different rates, creating stimuli with 152 distorted contents. In this work, an ACR methodology is used, and the point clouds rendered with *RColor* and the meshes with *RMesh* with color. The rendered videos were created and shown on a 24" LCD display. They show that meshes provide better quality at higher rates, while point clouds perform better for lower bitrate scenarios. They also compare V-PCC and G-PCC compression efficiencies using BD-MOS to show that V-PCC always yields better performance.

In [52], Wu *et al.* perform a subjective assessment study with 20 different point clouds from MPEG, JPEG and Sketchfab repositories. The V-PCC codec is used to code the point clouds for five different rate-distortion points. The DSIS methodology is used, and the subjects visualize meshes reconstructed from point clouds, rendered with *RMesh* with color interactively, with an HMD device. They show that in a VR environment, texture has a greater impact than geometry on the overall point cloud quality.

In [53], Su *et al.* apply several types of degradation (Gaussian noise, octree coding, G-PCC, and V-PCC coding) on a large point cloud dataset with different types of content. Each point cloud was created using a 3D reconstruction method receiving as input several views of an object. Some video sequences are generated from point clouds rendered with *RColor* and the DSIS methodology is used for the subjective assessment. It is shown that the V-PCC codec outperforms the other codecs, especially at low rates.

In [54], the subjective quality of dynamic encoded point clouds is assessed in an adaptive streaming scenario. Three point clouds are coded at five different rates using the V-PCC codec. The point clouds are rendered with *RColor* and visualized passively, using three different camera paths. Video sequences are shown to the users through an emulated network of three bandwidths, two different bit rate allocation schemes (view-focused and uniform) and two different viewport predictions (most recent and clairvoyant). It is shown that overall rating for volumetric video is lower than for traditional HD or 4K videos. Volumetric video is a 2D sequence of a dynamic point cloud. Regarding the objective-subjective correlation performance, they use 2D quality metrics for streamed sequences and they show that metrics need to be adjusted and scaled to properly match the human perception. They also show that viewport prediction leads to better results.

In [55], Subramanyam *et al.* perform a subjective study to evaluate the point cloud quality in 3-DoF and 6-DoF using a VR head-mounted device. Point clouds are encoded with MPEG V-PCC and the octree-based codec in [35] at four different rates. The subjective test follows an ACR methodology with Hidden Reference (HR) and the subjects could interact with the decoded point clouds rendered with *RColor*. Using the subjective study results, they show that V-PCC is significantly better than [35] for low rates while for higher rates the difference is statistically insignificant.

In [56], Perry et al. perform a subjective study to evaluate the MPEG G-PCC and V-PCC Intra codecs using static point clouds. The subjective test is conducted in four different laboratories using *RColor* rendering and following a DSIS methodology; a passive approach is used where video sequences are generated for predefined paths. The main conclusions regard the performance of objective quality metrics that will be presented in the next section.

This section presented a brief overview of the subjective quality assessment studies available in the literature. The summary of all these studies is presented in Table 2.1. The subjective studies conducted in this Thesis are also included in the table, highlighted in green. "Test Materials" regards the data that is used in the tests, "Test conditions" indicate the type of distortion addressed and the data used. Subjective test methodologies are listed under "Methodology". "Rendering" regards the point cloud rendering approach, the rendering primitives and whether color is used for rendering or not. The visualization device and the type of subjects' interaction with the point clouds are addressed in "Visualization" and "Interaction".

**Table 2.1:** Summary of the subjective point cloud quality assessment studies in the literature.

| Authors(s) | Year | Test Materials | | Test Condition | | Methodology | Rendering | | Visualization | Interaction |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Data | Type | Evaluation Domain | Degradations | | Approach | Primitives | | |
| Zhang *et al.* [34] | 2014 | Objects | Static | Geometry<br>Color | Down-sampling<br>Noise | Unspecified | *RColor* | Min-size Square | 2D Monitor | ✓ |
| Mekuria et al. [35] | 2017 | People | Dynamic | Geometry<br>Color | Octree Coding<br>JPEG Coding | ACR | *RColor w shading*<br>*RMesh* w Color | Min-size Square | 2D Monitor | ✓ |
| Mekuria et al. [36] | 2017 | People Objects | Static Dynamic | Geometry<br>Color | Octree Coding<br>JPEG Coding | ACR | *RColor* | Fix-size Cube | 2D Monitor | ✗ |
| Alexiou *et al.* [5] | 2017 | Objects | Static | Geometry | Gaussian Noise<br>Octree Coding | DSIS | *RPoint* wo Shading | Min-size Square | AR<br>2D Phone | ✓ |
| Alexiou *et al.* [6] | 2017 | Objects | Static | Geometry | Gaussian Noise<br>Octree Coding | DSIS | *RPoint* wo Shading | Min-size Square | 2D Monitor | ✓ |
| Alexiou *et al.* [37] | 2017 | Objects | Static | Geometry | Gaussian Noise<br>Octree Coding | DSIS<br>ACR | *RPoint* wo Shading | Min-size Square | 2D Monitor | ✓ |
| Alexiou *et al.* [8] | 2018 | Objects | Static | Geometry | Octree Coding | DSIS | *RMesh* | - | 2D Monitor | ✗ |
| Alexiou *et al.* [38] | 2018 | Objects | Static | Geometry<br>Visualization | Gaussian Noise<br>Octree Coding | DSIS | *RPoint* wo Shading | Min-size Square | 2D Monitor<br>2D Phone<br>AR | ✓ |
| Dumic *et al.* [39] | 2021 | People Objects | Static | Visualization | Not Distorted | ACR | *RPoint*<br>*RColor*<br>*RMesh* | Min-size Square<br>- | 2D Monitor<br>3D Monitor | ✗ |

| Authors(s) | Year | Test Materials | | Test Condition | | Methodology | Rendering | | Visualization | Interaction |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Data | Type | Evaluation Domain | Degradations | | Approach | Primitives | | |
| | | | | | | | *RMesh* w Color | | | |
| Alexiou *et al.* [41] | 2018 | Objects | Static | Geometry | Octree Coding | DSIS | *RMesh* | - | 2D Monitor | ✗ |
| Alexiou *et al.* [42] | 2018 | Objects | Static | Geometry | Gaussian Noise<br>Octree Coding | DSIS<br>ACR | *RPoint* wo Shading | Min-size Square | 2D Monitor | ✓ |
| Torlig *et al.* [44] | 2018 | People Objects | Static | Geometry<br>Color | Octree Coding | DSIS | *RColor* | Fixe-size Cube | 2D Monitor | ✓ |
| Alexiou *et al.* [45] | 2019 | People Objects | Static | Geometry<br>Color | Octree Coding | DSIS | *RColor* | Adaptive-size Cube | 2D Monitor | ✓ |
| Christaki *et al.* [43] | 2019 | People Objects | Static | Geometry | Mesh Coding | Nonstandard Pairwise | *RMesh* | Mesh | VR | ✓ |
| Alexiou *et al.* [47] | 2019 | People Objects | Static | Geometry<br>Color | G-PCC<br>V-PCC | DSIS | *RColor* | Adaptive-size Quads | 2D Monitor | ✓ |
| Zerman *et al.* [48] | 2019 | People | Dynamic | Geometry<br>Color | Down-sampling<br>V-PCC | DSIS<br>PWC | *RColor* | Fix-size Splat | 2D Monitor | ✗ |
| Zerman *et al.* [49] | 2020 | People | Dynamic | Geometry<br>Color | G-PCC<br>V-PCC | DSIS | *RColor* | Fix-size Square<br>- | 2D Monitor | ✗ |
| | | | | | Draco | | *RMesh* w Color | | | |
| Wu *et al.* [52] | 2020 | People Objects | Static | Geometry<br>Color | V-PCC | DSIS | *RMesh* w Color | Mesh | VR | ✓ |
| Su et al. [53] | 2019 | Objects | Static | Geometry<br>Color | Gaussian Noise<br>Octree Coding<br>G-PCC<br>V-PCC | DSIS | *RColor* | Min-size Square | 2D Monitor | ✗ |
| Van der hooft *et al.* [54] | 2020 | People | Dynamic | Geometry<br>Color | V-PCC<br>Streaming | ACR | *RColor* | Fix-size Quads | 2D Monitor | ✗ |
| Subramanyam *et al.* [55] | 2020 | People | Dynamic | Geometry<br>Color | V-PCC<br>Codec in [35] | ACR-HR | *RColor* | Fix-size Quads | VR | ✓ |
| Perry *et al.* [56] | 2020 | People | Static | Geometry<br>Color | V-PCC | DSIS | *RColor* | Fix-size Square | 2D Monitor | ✗ |
| Javaheri in Chapter 3, Impact of Denoising Algorithms | 2017 | Objects | Static | Geometry | Gaussian Noise | DSIS | *RMesh* | - | 2D Monitor | ✗ |
| Javaheri in Chapter 4, Impact of Coding | 2017 | People Objects | Static | Geometry | Octree Coding<br>Graph-based | DSIS | *RColor* | Adaptive-size Cube | 2D Monitor | ✗ |
| Javaheri in Chapter 5, Impact of Rendering | 2019 | People Objects | Static | Geometry | Octree Coding<br>V-PCC<br>G-PCC | DSIS | *RColor*<br>*RMesh*<br>*RPoint* | Fix-size Square<br>- | 2D Monitor | ✗ |

## 2.6 Point Cloud Objective Quality Assessment

This section will review the most relevant works in the literature on point cloud objective quality assessment after introducing some basic concepts.

### 2.6.1 Background and Key Concepts

As described in Chapter 1, objective quality assessment is essential for many point cloud applications. Point cloud quality metrics are much needed during the process of designing a point cloud codec, in the monitoring of the quality offered to users by point cloud communication

systems and for the optimization of modules in the point cloud pipeline shown in Section 2.2. Objective quality metrics can be classified into three categories:

- **Full reference (FR) quality metric:** The quality score is computed by comparing the reference and degraded point clouds.
- **Reduced reference (RF) quality metrics:** The quality score is computed by comparing some compact representation extracted from the reference and degraded point clouds.
- **No-reference (NR) quality metrics:** The quality score is computed only from the degraded point cloud and does not require the reference point cloud.

A point cloud is defined as a set of points, the geometry, with their associated attributes. Color (or luminance) information is the most important attribute, giving texture to the geometry, to create more realistic impressions of the object or visual scene. Both color and geometry may suffer degradations. In this context, point cloud quality metrics can be divided into two classes:

- **Point cloud geometry quality metrics:** These metrics only consider differences between the geometries (3D point coordinates) of the reference and degraded point clouds.
- **Point cloud geometry and texture quality metrics:** These metrics jointly evaluate the geometry and texture degradations. Note that attributes such as color can only exist associated to the points and thus these metrics consider the geometry in an implicit (only explicitly measuring texture distortions but using geometry along the process) or explicit (measuring geometry and texture distortions separately and performing some type of fusion) way.

Reference and degraded point clouds may be compared in the 3D world or mapped to 2D planes and compared as images or video sequences (e.g., obtained from multiple views). Based on the type of error (or similarity) that is computed from the reference and degraded point clouds, objective metrics (geometry and texture) can be classified into three groups:

- **Point-based metrics:** Compare the geometry or texture by establishing pointwise correspondences between the reference and degraded point clouds and measuring the distance (or positional error) between corresponding points. These quality metrics are usually computed in both directions (symmetric), i.e., between reference and degraded and vice-versa. Since point clouds are unordered, the corresponding point in the other point cloud is often computed based on nearest neighboring criterion. The D1-PSNR and D2-PSNR (defined later) are examples of these metrics and will be introduced in Section 0.
- **Feature-based metrics:** Extract some geometry and/or attributes features from each point cloud in a global or local way and use this intermediate feature-based representation to compute the error (distortion) between the reference and degraded point clouds. These metrics often need to use geometry information to associate local features between the two point clouds. However, when global features are used, such as color attributes, no geometry information is required.

- **Projection-based metrics:** Map 3D points into 2D planes and the error (distortion) is obtained by comparing the resulting 2D images, typically using a 2D quality metric. These metrics can evaluate texture and geometry (implicitly) together since the projected color is highly influenced by the geometry deformations.

### 2.6.2 Point Cloud Objective Quality Assessment: State-of-the-Art Review

In this section, the state-of-the-art objective quality assessment metrics are reviewed, following the classification proposed in Section 2.6.1.

**Point-based Point Cloud Quality Metrics**

The most popular point-based quality metrics for geometry are the point-to-point (Po2Po) [57] and point-to-plane (Po2Pl) [58] metrics. For every point in a degraded/reference point cloud, the nearest neighbor is obtained in the reference/degraded point cloud and the Haussdorff distance or the mean squared error (MSE) based distance is computed over all pairs of points; this type of metrics is usually referred as Po2Po metrics. The main disadvantage of this type of metrics is that they do not consider the fact that point cloud points represent the surface of an object(s) in the visual scene. To solve this issue, Point-to-Plane (Po2Pl) metrics have been proposed by Tian *et al.* [58], which model the underlying surface at each point as a plane; this plane is perpendicular to the normal vector at that point. Again, the point-to-point distance from every point to its nearest neighbor in the other point cloud is computed, which is then projected along the corresponding normal vector. This type of metrics results into smaller errors for the points that are closer to the surface. Currently, the MPEG-adopted point cloud geometry quality metrics are Po2Po MSE (D1) and Po2Pl MSE (D2) distances and their associated PSNRs [2]. These metrics are presented in detail in the next section.

The so-called Plane-to-Plane (Pl2Pl) metric proposed by Alexiou *et al.* [59], measures the similarity between the underlying surfaces associated to the points in the reference and degraded point clouds. In this case, tangent planes are estimated for both the reference and degraded points. As for Po2Pl metrics, tangent planes are used as a local linear approximation of the underlying object surface but, in this case, planes are estimated for both the reference and degraded point clouds. These metrics are affected by the difficulty to obtain reliable normal vectors for the decoded point clouds, especially when some types of compression artefacts are present (e.g., holes) or when the decoded point cloud is rather sparse [59]. This metric will be presented with more detail in the next section.

Although less popular, another type of metrics is the Point-to-Surface (or Point-to-Mesh) metrics [60]. In this case, a polygonal mesh is created from the reference point cloud and then the distance of each decoded point to the surface at the corresponding reference point is computed, thus considering the underlying surface (not necessarily a plane anymore). Point-to-Surface metrics are very much dependent on the specific surface reconstruction process and are better suited for mesh quality assessment.

There are not many point-based metrics considering color. However, the Po2Po PSNR of color in the $YC_bC_r$ color space is widely used by MPEG and in the literature to evaluate point cloud texture degradations. This metric works similar to Po2Po geometry metrics, but the error does not correspond to the distance of each point to its nearest neighbor. Instead, the difference between the color of the points in some correspondence is used, i.e., the error between the color of a point in one point cloud and the color of its nearest neighbor in another point cloud. This metric may also be computed only for the luminance (Y-PSNR) or chroma components ($C_b$/$C_r$-PSNR) individually, or as a weighted average of all color components.

**Feature-based Point Cloud Quality Metrics**

Most of the feature-based metrics, consider both the geometry and attributes, together. There is only one metric that only considers geometry features. In [61], Meynet *et al.* propose the PC-MSDM metric, a structural similarity-based point cloud geometry quality metric based on local curvature statistics. This metric computes the surface curvature associated to each point and then establishes point-based correspondences. The metric score corresponds to the Gaussian weighted curvature statistics for a set of local neighborhoods. Experiments show that this metric outperforms the Po2Po and Po2Pl MSE (D1 and D2) distance metrics for six point clouds in the subjective study presented in [37].

In [62], Viola *et al.* design a point cloud quality metric based on the histogram and correlogram of the luminance component. A correlogram allows to characterize the relationship between each pair of numeric variables in a dataset. Due to the unordered nature of point clouds, a specific color correlogram based on the k-nearest neighbor distance is proposed. The proposed point cloud quality metric shows better performance than Po2Po Y-PSNR Additionally, in this work, the proposed quality metric is fused with the Po2Pl MSE geometry metric (D2) using a linear model with a weighting parameter found using a grid search method. Experimental results show a significant improvement in the correlation performance of the final quality metric after fusion, compared to other point cloud quality metrics.

In [63], Diniz *et al.* propose the so-called Geotex metric which is based on Local Binary Pattern (LBP) descriptors adapted to point clouds and applied to the luminance. LBP is an efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considering the resulting binary number as a good feature for that point. To apply it on point clouds, the descriptor is computed on a local neighborhood corresponding to the k-nearest neighbors of each point in the other point cloud. Histograms of the extracted feature maps are obtained for both the reference and the degraded point clouds to be compared using a distance metric such as f-divergence [64] to compute the final quality score. Experimental results show that Geotex has a competitive and consistent performance comparing to other state-of-the-art quality metrics, notably Po2Po, Po2Pl and Pl2Pl metrics.

In [65], by Diniz *et al.* extend the Geotex metric by considering multiple distances, notably Po2Pl MSE for geometry and distance between LBP statistics from [63]. The quality score is the linear

combination of the Po2Pl MSE distance between points and their nearest neighbor in the other point cloud, Po2Po MSE distance between distance between LBP feature maps and the distance between the LBP feature map histograms, with the weights of this linear combination obtained using a least square error optimization. The results show that by including geometry distortions, the performance increases comparing to the original Geotex formulation.

In [66], Diniz *et al.* propose another quality metric which computes Local Luminance Patterns (LLP) on the k-nearest neighbors of each point on the other point cloud. The proposed metric includes several steps, such as voxelization, LLP descriptor creation, histogram distance computation and regression. The LLP descriptor associates a label to each target voxel considering the voxel luminance and the luminance of neighboring voxels. After, label histograms are computed, which are then compared with the Euclidean distance and mapped to the final score using some regression algorithm. The experimental results show that the logistic regression results in better performance than other regression methods. They also show that the proposed metric outperforms state-of-the-art metrics including several Po2Po and Po2Pl distance metrics and projection-based metrics in [44] and [47] for two different datasets.

In [67], Meynet *et al.* propose the Point Cloud Quality Metric (PCQM) metric, which combines the geometry features used in [61] with five color features related to lightness, chroma and hue. PCQM corresponds to the weighted average of the differences for geometry and color features between reference and decoded point clouds. The logistic regression is used to estimate the weights through cross-validation on a portion of the data. The metric is compared with D1-PSNR and D2-PSNR as well as Pl2Pl Mean Absolute Error (MAE) and outperforms these metrics significantly.

In [68], Viola *et al.* propose a reduced reference quality metric that jointly evaluates color and geometry. A set of seven statistical features such as mean, standard deviation, etc. are extracted from the reference and degraded point clouds in the geometry, texture, and normal vector domain, in a total of 21 features. The quality score is computed as the weighted average of the differences for all these features between the reference and degraded point clouds, with the weights obtained through a linear optimization algorithm. These algorithms maximize the Pearson Linear Correlation Coefficient (PLCC) between the objective and subjective scores obtained from a subjective study. The performance of the proposed metric is evaluated for four publicly available datasets to show the impact of the proposed features and the high performance of the metric.

Inspired by the SSIM quality metric for 2D images, Alexiou *et al.* proposes in [69] a quality metric using statistical local dispersion features. The features are extracted in a local neighborhood around each point in the reference and degraded point clouds considering four independent ‘attributes’, notably geometry, color (luminance), normal and curvature information. The metric is computed by pooling the differences of feature values between associated points in the reference and degraded point clouds, for each final feature corresponding to a combination of a dispersion feature and an ‘attribute’. The experimental results show that the best correlation happens considering only the variance of the luminance attribute, which outperforms all the state-of-the-art metrics.

**Projection-based Point Cloud Quality Metrics**

The first projection-based point cloud quality metric was proposed by Queiroz et al. in [70]; this metric was not compared with the alternative state-of-the-art metrics since it was mainly targeting to be used to codec rate-distortion optimization. This metric projects the reference and decoded point clouds onto the six faces of a cube around the point cloud, concatenates the corresponding images and measure the 2D PSNR of the resulting decoded image (using as reference the reference image).

In [44], Alexiou *et al*. designed a rendering software for visualization of point clouds on 2D screens, which was used to extract the orthographic projection of point clouds onto the six faces of a cube. A 2D quality metric is then applied to each image resulting from the projection and the final score is obtained with the average operator. In this way, the same content that is shown to subjective test subjects is also used for quality assessment by 2D quality metrics. Experimental results show that this metric outperforms Po2Po and Po2Pl MSE and Hausdorff as well as Pl2Pl MSE.

In [45], Alexiou *et al.* study the impact of the number of projected 2D images (each corresponding to a specific view) on the correlation performance of projection-based quality metrics. In this work, it is shown that even a single view could be enough to achieve high performance. Moreover, a projection-based quality metric weighting the images according to the user interactions performed during the subjective test is proposed. Experimental results show that the interactivity information obtained from users can be beneficial since the metric prediction power is significantly increased.

In [47], the quality metric proposed in [44] is benchmarked considering different number of views, different pooling functions, etc. The best performance of the projection-based metric is achieved when 2D quality metrics are applied on the projections from 42 different views and pooled with an l1-norm. However, the experimental results show that this projection-based metric cannot outperform the D1-PNSR and D2-PSNR geometry-only quality metrics.

In [71], a 2D quality metric is proposed to assess the point cloud quality, assuming that some 3D to 2D mapping already available can be used. In this case, the orthographic projections of the texture and depth are obtained for 6 views corresponding to the faces of a cube surrounding the point cloud. RGB to Gaussian Color Model (GCM) conversion is performed to use a color space more aligned with the human visual system. The depth-edge map is aggregated upon texture similarity between the reference and degraded point cloud as a local feature. Jensen Shannon (JS) for luminance is mathematically derived as the global feature. The contribution of each of the six projected images in local and global feature is weighted through a linear regression process and used in the final quality index between reference and degraded point clouds. Experimental results show that the proposed 2D quality metric designed for projected point clouds, outperforms the ten state-of-the-art 2D metrics used in the experiments.

The point cloud objective quality assessment metrics available in the literature are listed in Table 2.2 with their respective name, year, domain, and type. The quality metrics proposed in this Thesis are also included and highlighted in green for further comparison.

**Table 2.2:** Summary of point cloud objective quality assessment metrics in the literature.

| Author(s) | Metric Name | Year | Domain | Type |
|---|---|---|---|---|
| - | Po2Po MSE | - | Geometry | Point-based |
| MPEG 3DGC [57] | D1-PSNR | 2018 | Geometry | Point-based |
| Cignoni *et al.* [60] | Point-to-Surface | 1998 | Geometry | Point-based |
| Tian *et al.* | Po2Pl MSE | 2017 | Geometry | Point-based |
| MPEG 3DGC [57] | D2-PSNR | 2018 | Geometry | Point-based |
| - | Po2Po Hausdorff | - | Geometry | Point-based |
| Tian *et al.* [58] | Po2Pl Hausdorff | 2017 | Geometry | Point-based |
| Alexiou *et al.* [59] | Pl2Pl MSE | 2018 | Geometry | Point-based |
| Alexiou *et al.* [59] | Pl2Pl RMSE | 2018 | Geometry | Point-based |
| Meynet *et al.* [61] | PC-MSDM | 2019 | Geometry | Feature-based |
| Meynet *et al.* [67] | PCQM | 2020 | Geometry & Texture | Feature-based |
| Viola *et al.* [62] | $H_{L2}^{Y}$ | 2020 | Geometry & Texture | Feature-based |
| Viola *et al.* [62] | $d_{gc}$ | 2020 | Geometry & Texture | Feature-based |
| Viola *et al.* [68] | $PCM_{RR}$ | 2020 | Geometry & Texture | Feature-based |
| Alexiou *et al.* [69] | Point SSIM | 2020 | Geometry & Texture | Feature-based |
| Diniz *et al.* [63] | Geotex | 2020 | Geometry & Texture | Feature-based |
| Diniz *et al.* [66] | - | 2020 | Geometry & Texture | Feature-based |
| Queiroz et al. [70] | - | 2017 | Geometry & Texture | Projection-based |
| Torlig *et al.* [44] | - | 2018 | Geometry & Texture | Projection-based |
| Yang *et al.* [71] | - | 2020 | Geometry & Texture | Projection-based |

| Author(s) | Metric Name | Year | Domain | Type |
|---|---|---|---|---|
| Javaheri at Chapter 6, GH-PSNR | GH-PSNR | 2019 | Geometry | Point-based |
| Javaheri at Chapter 7, RA-PSNR | RA-PSNR | 2019 | Geometry | Point-based |
| Javaheri at Chapter 8, P2D Metric | P2D-MMD | 2020 | Geometry | Point-based |
| Javaheri at Chapter 9, Projection-based Metric | Pro | 2021 | Geometry & Texture | Projection-based |

### 2.6.3 MPEG and JPEG Point Cloud Objective Quality Metrics

Due to their importance in the point cloud objective quality assessment field, this section will review with more detail the metrics adopted in the MPEG and JPEG point cloud-related standardization activities. The Po2Po and Po2Pl PSNR metrics have been used in MPEG during the development of the G-PCC and V-PCC standard codecs and are known as D1-PSNR and D2-PSNR. Currently, JPEG also uses the Pl2Pl quality metric for its point cloud coding activities. These metrics serve often as the benchmarks for the novel, proposed metrics, thus somehow representing the state-of-the-art; due to their importance, they are explained in more detail in this section.

**D1-PSNR Quality Metric**

D1-PSNR is based on Po2Po MSE distance (called D1 in MPEG) which is computed, in a symmetric way, between the reference and degraded point clouds. In the following, the processes to compute the symmetric Po2Po MSE and Hausdorff metrics and after the D1-PSNR metric are presented. Po2Po objective quality metrics [57] establish point-wise correspondences in two directions, notably:

1) Direct $R \rightarrow T$: For each point in the reference (or original) point cloud *R*, the nearest neighbor (NN) points in the degraded (as test) point cloud *T* are identified.

2) Inverse $T \rightarrow R$: Correspondences are computed as for 1) but now in the opposite direction, thus, from point cloud *T* to point cloud *R*.

Assuming $\vec{e}_1(r_i, t_j)$ as an error vector between point $r_i$ in point cloud *R* and the $r_i$ nearest neighbor point $t_j$ in point cloud *T*, the Po2Point error vector length, i.e., the distance $d^{\text{Po2Point}}_{R,T}$ between these two points is given by:

$$d^{\text{Po2Po}}_{\mathbf{R,T}}(i) = \left\| \vec{e}\,(r_i, t_j) \right\|_2 \tag{2.1}$$

This distance is computed for all the points in both directions, i.e., from reference to degraded point clouds, $d^{\text{Po2Point}}_{\mathbf{R,T}}$, and from degraded to reference point clouds, $d^{\text{Po2Po}}_{\mathbf{T,R}}$. There are three main approaches to aggregate or pool all the computed errors:

- **Mean Squared Error (MSE):** Average of the squared distances between each point and their corresponding nearest neighbor in the other point cloud, for all points, as defined in:

$$\text{MSE}^{Po2Po}_{\mathbf{R,T}} = \frac{1}{N_R} \sum_{\forall r_i \in R} d^{\text{Po2Po}}_{\mathbf{R,T}}(i) \tag{2.2}$$

where $N_R$ is the number of points in the reference point cloud, $R$.

- **Hausdorff (HAUS) distance:** Maximum for all points of the MSE distances as defined in:

$$\text{HAUS}^{Po2Po} = \max_{r_i \in R}\{d^{\text{Po2Po}}(i)\}_{\mathbf{R,T}} \quad \mathbf{R,T} \tag{2.3}$$

- **Geometric PSNR:** Geometric PSNR metric defined as:

$$\text{PSNR}^{Po2Po}_{\mathbf{R,T}} = 10 \log_{10}\left(\frac{3P^2}{\text{MSE}^{Po2Po}_{\mathbf{R,T}}}\right) \text{ with } P = 2^{pr} - 1 \tag{2.4}$$

where $P$ is the peak constant value and $pr$ the point cloud coordinates precision. Since the metrics above are defined just for one direction (from point cloud $R$ to point cloud $T$), the metric needs to be also computed in the other direction in order the final symmetric metric score can be computed by pooling the maximum distance or minimum PSNR value. D1-PSNR is the pooled minimum of Po2Po PSNR in the two directions:

$$\text{D1} = MSE^{Po2Po} = \max\left(\text{MSE}^{Po2Po}_{\mathbf{R,T}}, \text{MSE}^{Po2Po}_{\mathbf{T,R}}\right) \tag{2.5}$$

$$HAUS^{Po2Po} = \max\left(\text{HAUS}^{Po2Po}_{\mathbf{R,T}}, \text{HAUS}^{Po2Po}_{\mathbf{T,R}}\right) \tag{2.6}$$

$$D1\text{-}PSNR = \min\left(\text{PSNR}^{Po2Po}_{\mathbf{R,T}}, \text{PSNR}^{Po2Po}_{\mathbf{T,R}}\right) \tag{2.7}$$

**D2-PSNR Quality Metric**

Po2Pl metrics [58] take into consideration the underlying object surface represented by a fitting plane to the local neighborhood of each point. Considering the 3D point locations and their associated surfaces, the normal for each point corresponds to the normal to the tangent plane at the surface. A point and the corresponding normal vector can, thus, determine the tangent plane for each point. As for Po2Po metrics, Po2Pl metrics are also symmetrically computed for both directions, i.e., from reference to degraded and from degraded to reference point clouds. However, Po2Pl metrics require the computation of normal vectors for the reference point cloud, which are used for the direct direction ($R \rightarrow T$). For the opposite direction ($T \rightarrow R$), the normal for each

point is estimated by averaging the normal vectors of the nearest neighbor points from the reference point cloud.

The Po2Pl error distance between two points $\vec{e}_2(r_i, t_j)$ in the reference and degraded point clouds is obtained by first computing the Po2Po error vector $\vec{e}_1$ which is then projected onto the normal $\vec{n}^{t}_{j}$. Thus, the Po2Pl distance $d^{\text{Po2Pl}}_{\mathbf{R,T}}(i)$ represents the error between a point and the corresponding plan/surface at the other point cloud as is given by:

$$d^{\text{Po2Pl}}_{\mathbf{R,T}}(i) = \left\| \vec{e}_2(r_i, t_j) \right\|^2 = \left( \vec{e}_1(r_i, t_j) \cdot \vec{n}^{t}_{j} \right)^2 \tag{2.8}$$

MSE distance, Hausdorff distance and Geometric PSNR may then be computed with the projected error distances as for Po2Po metrics in equations 2.2 to 2.4 (where the error vector is not projected). In this way, D2 and D2-PSNR come as:

$$\text{D2} = MSE^{Po2Pl}_{\mathbf{R,T}} = \max\left(\text{MSE}^{Po2Pl}_{\mathbf{T,R}}, \text{MSE}^{Po2Pl}_{\mathbf{R,T}}\right) \tag{2.9}$$

$$HAUS^{Po2Pl}_{\mathbf{R,T}} = \max\left(\text{HAUS}^{Po2Pl}_{\mathbf{T,R}}, \text{HAUS}^{Po2Pl}_{\mathbf{R,T}}\right) \tag{2.10}$$

$$D2\text{-}PSNR = \min\left(\text{PSNR}^{Po2Pl}_{\mathbf{T,R}}, \text{PSNR}^{Po2Pl}_{\mathbf{R,T}}\right) \tag{2.11}$$

In this way, the degraded point cloud points that are closer to the reference surface have smaller projected distances even though they are farther away from the corresponding point in the reference point cloud.

**Pl2Pl Quality Metric**:

This type of point cloud objective quality metric estimates the similarity of surfaces in the reference and degraded point clouds [59]. In this case, tangent planes are estimated for both the reference and degraded points. As for Po2Pl metrics, tangent planes are used as a local linear approximation of the underlying object surface but, in this case, planes are estimated for both the reference and degraded point clouds.

Again, to compute Pl2Pl metrics, the nearest neighbor correspondences are computed in both directions. The Pl2Pl metrics depend on the angular similarity (or dissimilarity) between the planes, i.e., the angular difference between the local planes associated to the points in a correspondence. In this case, the so-called *cosine similarity measure*, $CS_i$ measuring the cosine of the angle between two vectors is used. Here the two vectors correspond to the normal vectors (perpendicular to the tangent planes) for the two points in correspondence in the point clouds *R* and *T*, thus obtaining:

$$CS(i) = \cos(\theta) = \frac{\vec{n}^{r}_{i} \cdot \vec{n}^{t}_{j}}{\left\| \vec{n}^{r}_{i} \right\|_2 \left\| \vec{n}^{t}_{j} \right\|_2} \tag{2.12}$$

where $\vec{n}^{r}_{i}$ and $\vec{n}^{t}_{j}$ are normals for points $r_i$ and $t_j$ in point clouds *R* and *T*, respectively. To compute the angular difference (or distance), $d^{\text{Pl2Pl}}_{\mathbf{RT}}$, the inverse cosine is used as follows:

$$d_{\mathbf{RT}}^{\text{P2Plane}}(i) = 1 - \frac{2\arccos(|cs(i)|)}{\pi} \tag{2.13}$$

After computing the angular difference for all the points in the reference point cloud, different strategies may be used for pooling, i.e., for aggregating the angular differences obtained for all points. In this case, three pooling strategies were defined:

$$\text{MAD}_{\text{RT}} = \frac{1}{N_R} \sum_{\forall r_i \in R} d_{\mathbf{R,T}}^{\text{P2P}}(i) \tag{2.14}$$

$$\text{MSAD}_{\text{RT}} = \frac{1}{N_R} \sum_{\forall r_i \in R} \left(d_{\mathbf{R,T}}^{\text{P2P}}(i)\right)^2 \tag{2.15}$$

$$\text{RMSAD}_{\text{RT}} = \sqrt{\text{MSAD}_{\text{RT}}} \tag{2.16}$$

MAD stands for mean angular difference, MSAD for mean squared angular difference and RMSAD for the square root of MSAD. As for the other types of metrics, (2.14) - (2.16) are performed symmetrically, this means in both the direct and inverse directions, and the minimum value is selected as the final (similarity) quality score.

## 2.7 **Final Remarks**

This chapter introduced the point cloud processing pipeline and reviewed the impact of each key module on the perceived quality of a point cloud. The most relevant point cloud coding and rendering approaches that are going to be frequently used in this Thesis were presented in detail.

Finally, point cloud quality assessment was reviewed both in terms of subjective quality assessment studies and point cloud objective quality assessment metrics since these are the benchmarks for the reference developments to be proposed in this Thesis.

# Part II. Subjective Quality and Associated Objective Metrics Assessment

# Chapter 3

# Impact of Denoising Algorithms on Point Cloud Perceived Quality

## 3.1 Context and Objectives

Acquisition or creation is the first step of the point cloud processing pipeline shown in Section 2.2. Typically, point clouds represent the external surface of real or artificial objects. For real objects and visual scenes, point cloud data is directly or indirectly acquired by some sensors or is computationally created, e.g., via synthetic mesh-based models. There are nowadays several practical ways to acquire a point cloud, e.g., using an array of standard 2D cameras and some 3D reconstruction technique or, alternatively, some depth sensing devices such as Light Detection and Ranging (LiDAR) and Time-of-Flight Cameras. A rather popular consumer device is the Kinect sensor but, recently, also several LiDAR sensors are employed in autonomous cars to capture the surrounding environment as a point cloud. Both the point clouds generated from multiple images using 3D reconstruction techniques and those obtained with depth sensors, like the Kinect structured infrared light approach, are often rather noisy [72]. The noise in point clouds poses great challenges at the coding stage but also at the subsequent rendering (or surface reconstruction) stage [73]. Thus, point cloud denoising is an essential step to enable a more faithful rendering of the visual scene and the efficient compression of the point cloud data. In fact, in many application scenarios, point cloud noise may lead to significant bitrate costs and lower the quality of experience for the final users. Considering the increasing importance of the point cloud data 3D representation model, research around point cloud denoising is becoming increasingly important in order better experiences are offered to the users.

In this context, the reliable quality assessment of point cloud data is fundamental to evaluate the performance of the several steps in a point cloud processing pipeline, notably acquisition,

denoising, coding, streaming and so on. However, because subjective quality assessment cannot be performed too often, the reliable quality assessment of the point cloud data quality asks for appropriate objective quality metrics as they can provide critical useful guidance for the design of point cloud-based systems. In this context, this chapter addresses two main objectives:

- **Design of a subjective assessment methodology for point cloud geometry denoising algorithms** assuming arbitrary and unstructured point clouds, without any underlying connectivity. This type of methodology is rather important since point clouds are rendered before being displayed and thus the presence of noise can lead to annoying artifacts, which lead to lower perceived quality. To achieve this objective, classical denoising solutions were adapted to either remove erroneous points or improve its positioning.
- **Study of the correlation performance of available objective quality metrics to evaluate the quality of denoised point cloud geometries.** Since reliable objective quality metrics are critically needed, this study targets evaluating the objective-subjective correlation of available objective quality metrics for point clouds which have been denoised. It is important to keep in mind that the perceived quality of a (point cloud) surface is mostly related to the positioning of the points in relation to the original surface, even if their position is changed.

To reach these objectives, this chapter is organized as follows. Section 3.2 describes the point cloud denoising approaches used in this chapter. Section 3.3 describes the subjective assessment study performed and used in correlation performance studies. Section 3.4 presents the experimental results and their analysis. Section 3.4.1 evaluates the point cloud denoising algorithms while Section 3.4.2 brings the correlation performance of the state-of-the-art metrics at time of this work. Finally, Section 3.5 concludes this study on impact of denoising algorithms on perceived quality of the point cloud.

## 3.2 Point Cloud Denoising Approaches

In practice, there are two distinct ways to generate the point cloud geometry data: by direct acquisition, where the geometry is obtained using sensors which measure the distance to the objects/scene, and by indirect acquisition, where the geometry is obtained by means of 3D reconstruction from a set of different perspectives (views) from the scene. In both cases, several errors can occur, and point clouds are usually noisy. In the first case, sensors usually have limitations and make measurement errors due to environmental issues such as illumination, materials reflectance, and imperfect optics. In the second case, the wrong estimation of disparity (feature correspondences), imprecise depth triangulation, and inaccurate camera parameters, can lead to significant geometry errors. The geometry errors corrupting a point cloud can be classified in two types: outlier errors and positioning errors. The outlier errors occur when there are some points which are far away from the underlying object surface that should be represented. These errors are very common in the indirect acquisition case when the disparity estimation between views fails. The positioning errors occur when the points are only slightly out of their correct

position. This typically occurs when there are imprecisions in the depth estimation (e.g., 3D triangulation for indirect acquisition) and due to sensor noise (direct acquisition). Typically, position errors are modelled with white Gaussian noise while the outlier errors are better modelled with impulse noise.

The denoising of point clouds is a difficult problem since there is not any connectivity information between points and therefore surface topology is not directly represented. Most of the point cloud denoising algorithms estimate the underlying surface, which is naturally more difficult or less accurate when outliers are present. Thus, the first step is to identify and remove outliers before the point cloud data is further processed since outliers do not contain relevant information about the underlying surface and can reduce the performance of any position denoising algorithm. The next sections review outlier removal and graph-based positioning denoising algorithms.

### 3.2.1 Outlier Removal Algorithms

The first step in point cloud denoising is to remove the points which are very far from the underlying surface, which they should represent. Most of the outlier removal algorithms perform, for each point, a statistical analysis using a set of its neighboring points and apply some criteria to decide if the point under test is an outlier or not. The following two solutions for outlier removal are described as examples:

- **Radius outlier removal** [74]**:** This algorithm exploits the fact that outlier points are typically far away from its neighbors. Thus, in this method, every point is connected to its neighbors in a small ε radius using a small graph. The points corresponding to outliers have a significantly smaller number of neighbors than the other points, i.e., the number of graph nodes is smaller. The outliers can be identified by establishing a threshold τ on the minimum number of neighbors in the $\varepsilon$ radius neighborhood.
- **Sparse outlier removal** [75]: This algorithm also exploits the distance from each point to the set of its neighboring points. In this case, it is assumed that the distribution of the distance between each point and its neighboring points is a normal distribution. For each point, the mean distance, $u$, and the standard deviation, $\sigma$, to all its neighbors are calculated. All the points that are out of a confidence interval $u \pm \alpha\sigma$, defined from this distribution, are considered as outliers. The parameter α should depend on the size of the analyzed neighborhood.

Naturally, the points detected as outliers are removed from the point cloud. It is assumed that the number of outliers is not typically large as otherwise this process may create holes or insufficient number of points in some regions and, thus, reduce the point cloud quality. The two outlier removal methods presented above are available in the popular Point Cloud Library (PCL) [15].

### 3.2.2 Position Denoising Algorithms

To After outlier removal, the remaining 'wrong' positions should be corrected using some local regularization criteria. In this case, the distribution of the point positions has usually a certain degree of smoothness or regularity since these positions describe a 3D object/surface in space. To perform position denoising, a *K-NN* graph $G$ is constructed to represent the underlying object surface; here the spatial coordinates are considered as signals residing on each node of this graph. The creation of a *K-NN* graph is a simple procedure, where each point is connected to its *K* nearest neighbors and then any duplicated connections are removed. Each connection established is regarded as an edge and the Euclidean distance is used to measure the distance between points. Since the graph represents the geometric structure defining the shape of an object, it is now only necessary to associate the 3D coordinates of each point to each graph vertex.

At this stage, to perform denoising, graph-based equivalents of classical denoising algorithms [74] can be applied. This means the signal denoising problem on a graph may be addressed as a convex optimization problem with the constraint that the denoised signal must be smooth on the graph as for ordinary signals [76]. There are in practice several ways to regularize the signals on the graphs, in this case the 3D positions, notably:

- **Tikhonov regularization** - The most straightforward method is to apply a *Tikhonov (TK)* regularization term using the $L_2$ norm where the graph denoising corresponds to solving the following optimization problem:

$$\dot{x} = \arg\min_x \|x - f\|_2^2 + \gamma \|\nabla_G x\|_2^2 \tag{3.1}$$

where $\dot{x}$ is the estimation of the denoised signal, $f$ is the noisy signal, $\gamma$ is the regularization parameter and $\nabla_G x$ is the gradient of the signal on graph *G* [74] [77]; however, other regularization terms may be applied.

- **Total variation regularization** - When a *Total variation (TV)* regularization term is used, the following optimization problem must be solved:

$$\dot{x} = \arg\min_x \|x - f\|_2^2 + \gamma \|\nabla_G x\|_1 \tag{3.2}$$

These two optimization problems can be solved using the Alternating Direction Method of Multipliers (ADMM) [78]. The denoising algorithms were implemented with GSPBox [79] for graph signal processing and UNLocBoX [80] for convex optimization.

## 3.3 Subjective Assessment Framework

In this section, the subjective assessment setup and the creation of the video content for the subjective test are described.

### 3.3.1 From Point Clouds to Video Sequence

The *Bunny*, *Armadillo*, *Dragon* and *Happy Buddha* point clouds from the Stanford 3D scanning repository [81] were selected; renderings of these PCs are shown in Figure 3.1. At the time of this work, voxelized point clouds were not available and, thus, the point clouds in this dataset have floating point precision. To perform a fair comparison, the point cloud coordinates have been normalized to a dynamic range between [0,1] and sub-sampled to have a similar number of points, in this case around 50000; this should allow the point clouds to have rather similar density. The voxel grid down-sampling method from the PCL library was used to sub-sample the point clouds to the target resolution. High-quality point clouds have been selected for this experiment in order a ground truth without noise is available. After, the same type and amount of noise is added to all point clouds to achieve an equivalent subjective quality degradation. The noisy point clouds are then processed by the outlier removal and position denoising algorithms described in Section 3.2. For the outlier removal algorithm, the radius was set to 0.01 and the threshold $\tau$ to 3. The regularization parameters for the position denoising algorithms were experimentally defined as $100 \times \sigma_{noise}^{2}$.

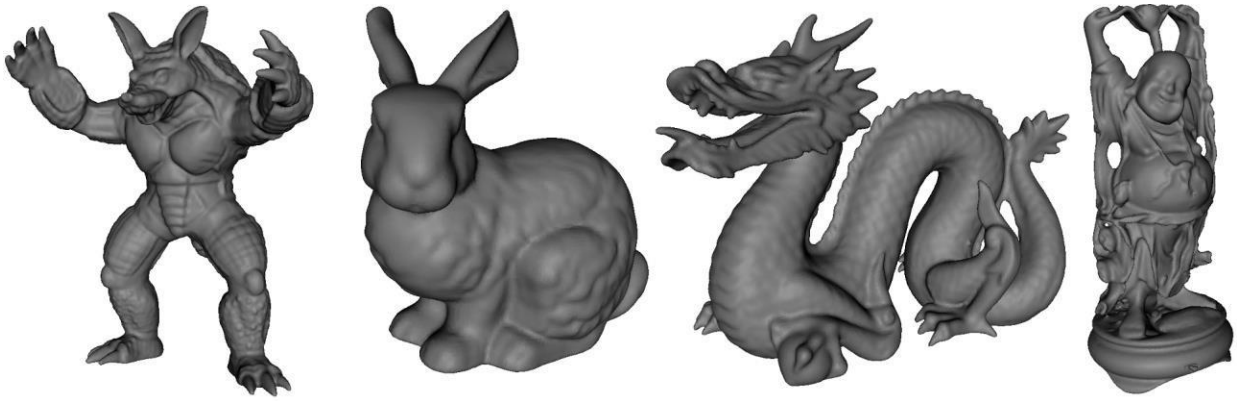


**Figure 3.1:** Stanford repository point clouds selected for the denoising quality study: *Armadillo*, *Bunny*, *Dragon*, and *Happy Buddha*, rendered as a mesh with mid gray color.

After, the denoised point clouds are further processed to create a rendered video to be visualized by the final user, in this case in a standard 2D display. This processing should allow the test subjects to evaluate the point cloud quality in a rather realistic way, notably observing the impairments on the object surface from multiple viewing directions. The processing of the denoised point clouds towards the subject visualization involves the following steps:

1. Poisson surface reconstruction [82] is used to create a 3D surface of the object; this process was performed using the Cloud Compare software [83].
2. To allow the users to visualize the denoised point clouds from multiple perspectives, a set of relevant viewpoints is defined and a video with smooth transitions between viewpoints is created, i.e., the motion of a virtual camera following a pre-defined path between viewpoints

is simulated. In this case, the point cloud is first rotated around a vertical axis and after rotated around a horizontal axis.

The final sequence with all viewpoints is stored as a high-quality video sequence encoded in a visually lossless way, with a 1600×800 spatial resolution, 25 fps, with a total duration of 10 seconds.

### 3.3.2 Subjective Assessment Methodology

The DSIS assessment methodology was adopted to obtain the subjective scores from the subjects in this experiment. All the subjects were selected from *Instituto Superior Técnico (IST)* professors and MSc and PhD students, including 4 females and 17 males; only 4 subjects were familiar with the field of point cloud or mesh representation. The goal of the experiment was initially briefly explained to the subjects, and they were asked to participate in a short training session just before the test session. The point cloud reference and rendered videos were viewed and scored by each subject in a test session lasting around 30 minutes where the play of each new point cloud video sequence was controlled by the subjects as they had to explicitly press 'Play'. All rendered video sequences were organized into 9 rounds with each round including a video sequence for each test point cloud; the order of the point clouds was random within each round. Each round proceeded according to BT-500.13 recommendation [31] with the subjects watching first the reference rendered point cloud sequence (without noise degradation) and after the impaired/denoised video sequence for each point cloud pair. After, the second video impairment was scored with respect to the reference video in a 1-5 scale associated to five quality levels, notably 'very annoying', 'annoying', 'slightly annoying', 'perceptible but not annoying' and 'imperceptible' from 1 to 5. A so-called *IST Point Cloud Subjective Assessment Application* has been designed and developed using *MATLAB* and fulfilling the BT-500.13 specifications. Figure 3.2 shows the viewing and scoring panels from the used quality assessment application.

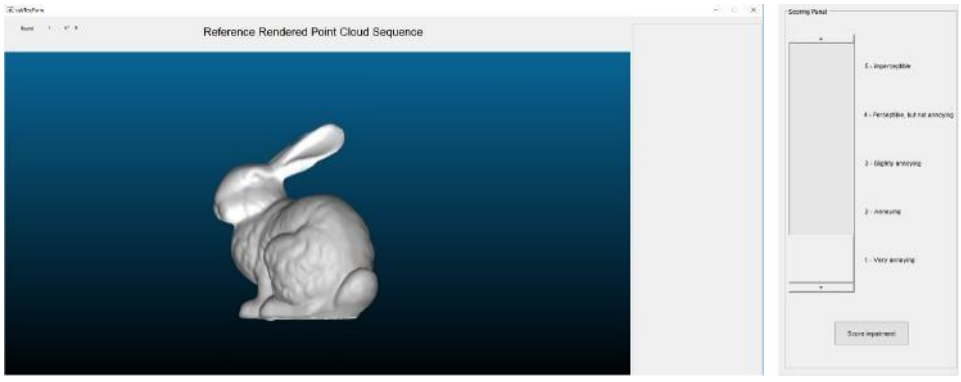



**Figure 3.2:** *IST Point Cloud Subjective Assessment Application*: scoring panel is shown only after viewing both the reference and degraded point cloud rendered videos pair.

The point cloud rendered video sequences were visualized on a 23-inch ASUS VH238 monitor with 1920×1080 resolution. An i7 workstation with the Intel HD 530 graphic card and 128MB video memory was used to play the test sequences at the correct frame rate without any delays.

### 3.3.3 Subjective Scores Processing

The same outlier removal process as described in Chapter 2 was deployed before further processing the subjective scores. In our study, none of the subjects was rejected. After outlier removal, the average of all scores across the subjects were computed for each point cloud stimulus, thus obtaining the overall MOS.

## 3.4 Performance Evaluation

This section focuses on the evaluation of the point cloud denoising algorithms with subjective and objective quality scores, as well as on the assessment on the correlation between the obtained objective and subjective scores. First, Section 3.4.1 reports the subjective performance of the outlier removal and position denoising algorithms. Then, Section 3.4.2 presents the subjective-objective correlation for various objective quality metrics, notably the most used metrics at the time this work was developed.

### 3.4.1 Denoising Algorithms Assessment

To study the impact of denoising algorithms on the point cloud perceived quality, first the outlier removal algorithms are objectively assessed, and the best performing outlier removal algorithm is selected. After, the subjective performance of the positioning denoising algorithms is assessed (after outlier removal).

**Outlier Removal Algorithms: Objective Assessment**

The outlier removal algorithms were first objectively evaluated to select the best one. In this case, Po2Po objective quality metrics were employed since it is difficult to reliably estimate the normal vectors for Po2Pl metrics when outliers are present. To simulate the presence of outliers, 5, 10 and 15 percent of the points in the reference point clouds for the four test point clouds were affected by impulse noise and were after denoised using the two algorithms presented in Section 3.2.2. The final degraded point clouds were after obtained by randomly selecting 5, 10, and 15 percent of their points and adding impulse noise to them. Impulse noise corresponds here to Gaussian noise, with zero mean and a respective large variance ($\sigma^2 = 1$ in experiments of this Thesis where contents are normalized between 0 and 1). The average results for the four test point clouds are shown in Table 3.1 for the Root Mean Squared Distance (RMSD) and Hausdorff distance Po2Po objective quality metrics; thus corresponds, in practice, to two types of data pooling from the same Po2Po distances, the average and the maximum.

The results in Table 3.1 show that the radius outlier removal algorithm can remove the outliers more efficiently in comparison with the sparse outlier removal algorithm, considering the RMSD

and Hausdorff quality metrics (except for 10% noise level). The Hausdorff metric seems to be less stable for this case, very likely because it is a maximum distance and, thus, a single outlier can lead to significant quality variations. Based on these results, it was decided to use the radius outlier removal algorithm for the rest of the experiments, i.e., as an outlier removal pre-processing step for the two position denoising algorithms.

**Table 3.1:** Objective quality assessment results for outlier removal algorithms.

| | Noise Level (%) | Po2Po Metrics | |
|---|---|---|---|
| | | ***RMSD*** | ***Hausdorff*** |
| ***Noisy Point Cloud*** | 5 | 19.47 | 426.578 |
| | 10 | 27.73 | 447.874 |
| | 15 | 33.85 | 548.341 |
| ***Denoised Point Cloud after Radius Outlier Removal*** | 5 | **0.818** | **19.477** |
| | 10 | **1.063** | 34.837 |
| | 15 | **1.271** | **31.795** |
| ***Denoised Point Cloud after Sparse Outlier Removal*** | 5 | 1.364 | 23.505 |
| | 10 | 2.618 | **29.140** |
| | 15 | 3.867 | 53.414 |

**Position Denoising Algorithms: Subjective Assessment**

The performance of the position denoising algorithms was subjectively evaluated for three different levels of Gaussian noise and four point clouds. Impulse noise was added as in the previous section and the remaining points were impaired by zero mean Gaussian noise with three different powers representing 'small' ($\sigma^2 = 0.003$), 'medium' ($\sigma^2 = 0.007$) and 'high' ($\sigma^2 = 0.01$) levels of noise. In this subjective assessment, the noisy point clouds were also included. In Figure 3.3, MOS values and the corresponding confidence intervals (defined as in [84]) are shown. The subjective results show that, for all the point clouds and both position denoising algorithms, the point cloud quality is improved with respect to the noisy versions. Naturally, the quality improvements are larger for the medium and high amounts of noise. Also, the improvements are larger for the simpler *Bunny* and *Armadillo* point clouds and lower for the more complex *Dragon* and *Happy Buddha* point clouds, thus suggesting that the shape and surface cloud regularities play an important role in the denoising process.

The *TK* and *TV* regularization functions lead to similar results with advantage for the latter in the *Armadillo* and *Happy Buddha* point clouds. However, *TK* has managed to outperform *TV* in some cases, despite the smoothness of the solution, i.e., even when some high frequency components are removed from the point cloud.

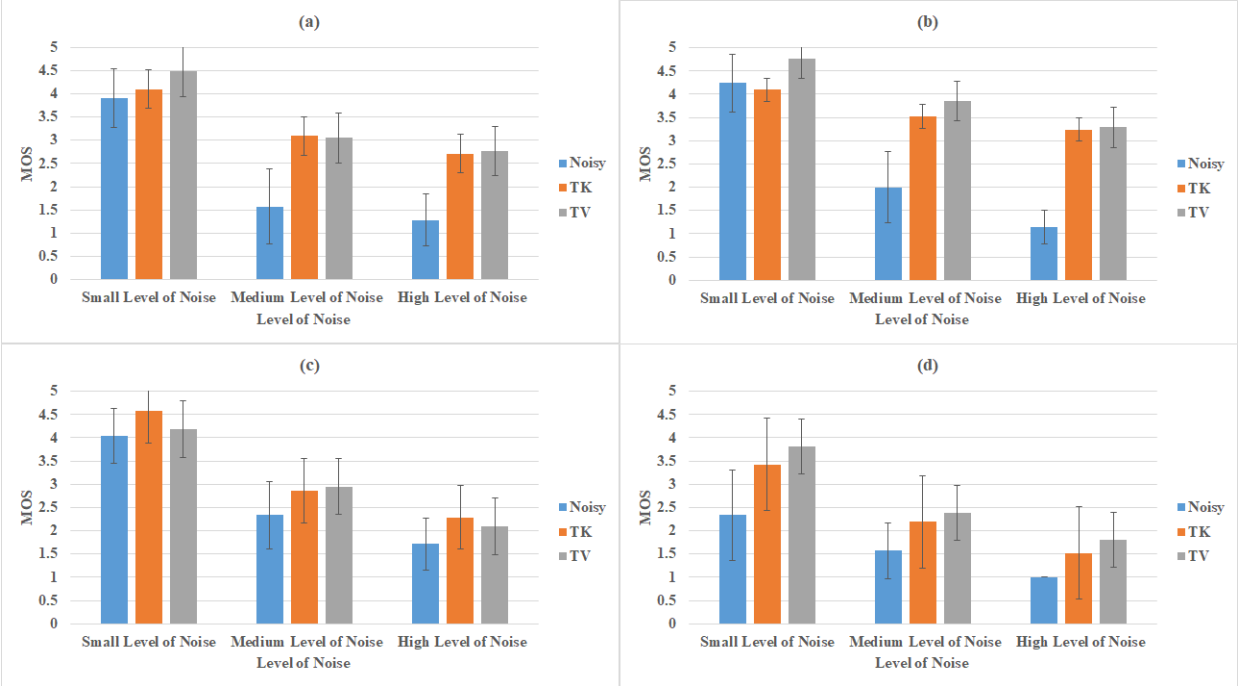



**Figure 3.3:** Mean opinion scores for the noisy and denoised point clouds for three different levels of noise: a) *Bunny*; b) *Armadillo*; c) *Dragon*; d) *Happy Buddha*.

### 3.4.2 Objective Quality Metrics Correlation Assessment

The subjective-objective correlation performance of the four selected objective quality metrics (Po2Po RMSD, Po2Po Hausdorff and Po2Pl RMSD, Po2Pl Hausdorff) has been evaluated using: i) Pearson Linear Correlation Coefficient (PLCC) as a measure of the linear dependence between the MOS and the corresponding objective metric scores; ii) Spearman Rank Order Correlation Coefficient (SROCC) as a measure of the strength and direction of monotonicity between the MOS and the corresponding objective metric scores; and iii) Root Mean Square Error (RMSE) as the average error between corresponding objective and subjective scores.

Before performance assessment, a nonlinear logistic function was applied on the objective quality scores to map them to the subjective score scale, as recommended in [32], thus obtaining the predicted subjective scores. Let $Q_i$ be a specific objective quality score for video $i$ and $\beta_k$ the logistic function parameters. According to [32], the objective quality score $Q_i$ for a specific objective metric is used to obtain a predicted subjective score $MOS^i_p$ for video $i$ according to:

$$MOS^i_p = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-\left(\frac{Q_i - \beta_3}{\beta_4}\right)}} \tag{3.4}$$

The logistic function parameters, $\beta_1$, $\beta_2$ and $\beta_3$, were estimated by performing a non-linear regression of the objective metric scores over the corresponding MOS scores. This allows fitting

the objective metric scores to the MOS scores. In this case, the initial estimates for the parameters in (3.4) were defined following the suggestions in [32] and the fitting was performed using the *MATLAB* function for non-linear regression. Figure 3.4 shows the predicted subjective scores (based on objective scores) versus the obtained MOS for all the point cloud test set used in this experiment.

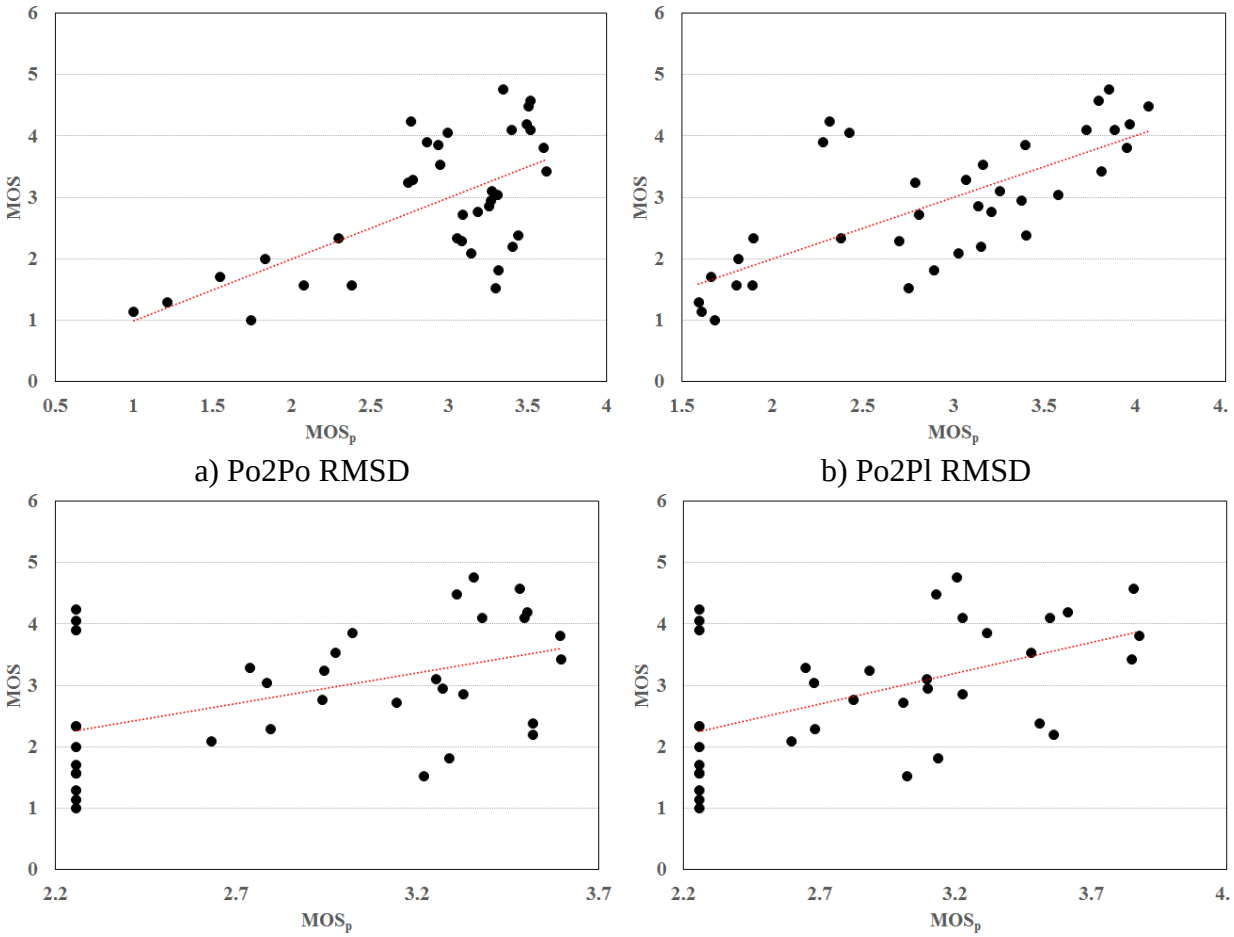



**Figure 3.4:** $MOS_p$ (predicted subjective scores) vs. MOS (real subjective scores) for the noisy and denoised versions of all point clouds.

The Po2Pl RMSD points are closer to the red line where $MOS = MOS_p$, corresponding to the optimal performance for an objective quality metric. However, for all the objective quality metrics, many points are still far away from the optimal correlation performance, clearly suggesting that there is still room for improved objective quality metrics, at least for this type of quality assessment task. Table 3.2 shows the correlation performance for the four selected objective quality metrics in terms of PLCC, SROCC and RMSE for the noisy, denoised and all (noisy + denoised) point clouds considering the four point clouds selected; the best correlation behaviors are highlighted in bold.

Correlation performance (PLCC, SROCC, and RMSE) for the three selected objective quality metrics.

**Table 3.2:** Correlation performance (PLCC, SROCC, and RMSE) for the three selected objective quality metrics.

| | | **PLCC** | | | **SROCC** | | | **RMSE** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Objective metrics | | Noisy | Denoised | All | Noisy | Denoised | All | Noisy | Denoised | All |
| RMSD | Po2Po | 0.87 | 0.58 | 0.64 | 0.78 | 0.51 | 0.55 | 0.55 | 0.72 | 0.82 |
| | Po2Pl | **0.89** | **0.81** | **0.73** | **0.85** | **0.79** | **0.72** | **0.51** | **0.51** | **0.73** |
| Hausdorff | Po2Po | 0.53 | 0.33 | 0.48 | 0.45 | 0.34 | 0.47 | 0.95 | 0.83 | 0.94 |
| | Po2Pl | 0.57 | 0.44 | 0.51 | 0.46 | 0.47 | 0.51 | 0.93 | 0.79 | 0.91 |

From the obtained subjective-objective correlation results, the following conclusions can be drawn:

- **Po2Po vs Po2Pl objective quality metrics**: For the denoised point clouds, the Po2Pl metrics have higher correlation with the subjective scores since these metrics model the underlying point cloud surface. Note that the typical distortions for the denoised point clouds correspond to slight displacements in the underlying surface, which may lead to significant errors for the Po2Po metrics, without a corresponding significant perceptual impact after rendering. Figure 3.5 shows the *Bunny* point cloud with high noise on the left and the corresponding denoised result after *TK* regularization on the right; in red, aligned with the reference point cloud in black. Interestingly, the Po2Po RMSD score for the noisy point cloud (0.0077) is lower than for the denoised point cloud (0.0080), which shows a cleaner surface; at the same time, Po2Pl RMSD is 0.0068 and 0.0058 for the noisy and denoised point clouds, respectively, which better expresses the associated perceptual quality.

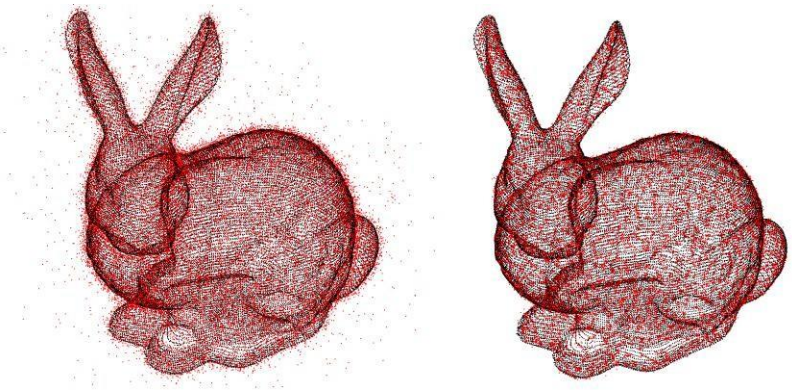


**Figure 3.5:** Noisy point cloud (left) and TK denoised point cloud (right) with the reference point cloud in black and the processed point cloud in red.

- **RMSD vs. Hausdorff distance-based objective quality metrics:** The Hausdorff-based metrics correlate poorly with the subjective scores both for the Po2Po and Po2Pl metric types. As the Hausdorff-based metric captures the maximum distance, it cannot capture well the perceptual distortion occurring in a full point cloud, especially for cases where there is a rather local high distortion. This result also seems to imply that some localized distortions are subjectively better tolerated. On the other hand, Po2Pl RMSD has the best correlation performance for noisy, denoised and all data, thus showing the power of error average pooling (and not maximum) in a similar way to traditional 2D image/video metrics.
- **Noisy vs. denoised point clouds:** For noisy point clouds, the Po2Po and Po2Pl metrics performances are rather similar, since the computation of normal vectors for Po2Pl metrics is less reliable and, thus, some errors may be introduced, reducing the positive impact of Po2Pl metrics. However, for the denoised point clouds, points are closer to the true object surface and, thus, the Po2Pl metrics correlation increases.

## 3.5 **Final Remarks**

The main objective of this chapter is to reliably assess point cloud position denoising algorithms through subjective tests as well as the subjective-objective correlation performance of the point cloud objective quality metrics that were available at the time of this work when applied in a denoising context. Experimental results show that Po2Pl metrics have better correlation with human perception when evaluating distortions from point cloud denoising algorithms. The RMSD is also much better correlated with human perception than the Hausdorff distance where localized distortions have a much higher impact. As future work, the objective quality metrics will be evaluated for decoded point cloud, where different types of distortions are present. This issue will be addressed in Chapter 4 and Chapter 5 along with the rendering impact.

The work presented in this chapter has been included in a conference paper published at Hot3D - IEEE International Conference on Multimedia & Expo Workshops:

**A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, “Subjective and objective quality evaluation of 3D point cloud denoising algorithms,” *International Conference on Multimedia & Expo Workshops*, Hong Kong, July 2017.

# Chapter 4

# Impact of Coding on Perceived Quality of Point Cloud

## 4.1 Context and Motivation

As mentioned in Chapter 1, efficient coding is essential to store and transmit point clouds; however, to reach higher compression factors and potentiate more applications, lossy coding is typically used, thus implying that some degradation is associated to the geometry and color of the decoded point cloud. To evaluate the point cloud quality after coding, objective quality metrics are typically used, since subjective quality assessment is rather time consuming and expensive. However, very few subjective quality assessment experiments with decoded point clouds had been performed at the time this work was performed. The only subjective quality assessment study on decoded point clouds only addressed the quality of decoded geometry using an octree-based point cloud codec [85]. For example, the assessment of the objective-subjective correlation of point cloud quality metrics was missing, despite its importance for the design and optimization of efficient point clouds codecs for the transmission and storage of point cloud data, namely considering the quality degradations introduced by recent coding schemes.

In this context, it was urgent to perform a solid assessment of the correlation performance of point cloud quality metrics while considering the most recent point cloud codecs and associated artefacts. To compare different point cloud geometry codecs, the coding rate and the decoded geometric quality/distortion need to be reliably evaluated. A reliable objective quality metric can help to find the best RD performance trade-off while designing a coding scheme for any type of media, thus improving the final Quality of Experience (QoE) for the end-users. In this context, the main objective of this work was:

- **To study the objective-subjective correlation of available point cloud objective quality metrics for geometry** and thus, evaluate their performance in assessing the geometry quality of decoded point clouds.

In this work, two popular point cloud coding solutions were used to code two point cloud datasets with rather different characteristics. Moreover, two types of objective quality metrics for the geometry component of point clouds were considered, notably Po2Po and Po2Pl metrics. A DSIS methodology was used for the subjective assessment. The subjective experiments conducted at *Instituto Superior Técnico (IST)* were designed with the aim to evaluate the selected objective point cloud geometry quality metrics. To the best of our knowledge, this was the first study on the performance of objective quality metrics with decoded geometry point cloud data. Therefore, the conclusions of this work were very important for the design of future point cloud codecs.

The rest of this chapter is organized as following: Section 4.2 describes the selected point cloud coding solutions. Section 4.3 describes the adopted subjective assessment framework. Next, Section 4.4 presents the experimental results obtained for the objective quality metrics correlation assessment and the associated key conclusions of this study. Finally, Section 4.5 summarizes and concludes this chapter.

## 4.2 Point Cloud Coding Approaches

Efficient coding is crucial for many applications that require storage on a limited capacity device or transmission through a network with a limited bandwidth; this is especially true for point cloud data which uncompressed size may be huge [86], [85], [87], and [88] . In this chapter, the key objective is to study the performance of full-reference objective quality metrics to measure the quality of decoded point cloud geometry with respect to the original point clouds geometry. Although, point cloud coding methods may be lossless or lossy, to transmit point clouds in resource-constrained networks (e.g., via wireless channels), it is critical to employ lossy coding to reduce the coding rate more effectively. In this chapter, two lossy geometry coding methods are selected to encode the points 3D locations and, therefore, to obtain decoded point clouds which are impaired by some coding artifacts. In this work, the objective is to assess the correlation performance of objective quality metrics just for point cloud geometry; in this context, attribute data such as the color of each point is left uncoded even if it may be used for rendering.

The first selected point cloud geometry coding solution is rather popular and based on the organization of all points in an octree structure, which nodes are after scanned to extract their occupancy. The popular point cloud octree-based codec that is available in the so-called Point Cloud Library (PCL) [89] was selected since it was a very popular codec at the time of this work, also considered by MPEG for the future developments of point cloud coding standards. The second selected point cloud geometry coding solution was designed and developed by the author of this Thesis and organizes all points in a graph structure, using the nearest neighbors to apply a graph transform. These two point cloud geometry coding solutions introduce different types of artifacts: while the octree-based coding solution creates some type of down-sampling effect in the decoded

point cloud geometry, the graph-based coding solution introduces blocking artifacts which are also typical in frequency-domain DCT-based image codecs.

### 4.2.1 Octree-based Geometry Coding

This coding solution exploits the octree structure to encode the $(x, y, z)$ 3D coordinates data (Chapter 2, Section 2.3.1). In an octree, each node is referred as a voxel and represents a cube in the 3D space. The root node is a cube containing all points of the point cloud (also referred as bounding box). Then, starting from the root node, this cube is divided into 8 cubes with the same size, which correspond to each root node child. These cubes correspond to the voxels. This process is repeated for every single child node, iteratively, which are again divided into 8 nodes. Naturally, a tree node is not divided if there is no point occupying the corresponding voxel. The decoded quality is determined by the number of times this process occurs, i.e., by specifying the octree depth and, therefore, indirectly limiting the minimum voxel size; another possibility is to directly define the minimum voxel size. When the point cloud is decoded, all the points inside a voxel (that is occupied with 1 or more points) are represented with just one point at the center of the voxel. The occupancy of a node can be represented with a single occupancy byte, indicating which child nodes (leaf voxels) are occupied, and which are not. By traversing the tree in breadth-first order and outputting the corresponding occupancy byte for each voxel, the point cloud geometry can be efficiently compressed. Figure 4.1 illustrates this process.

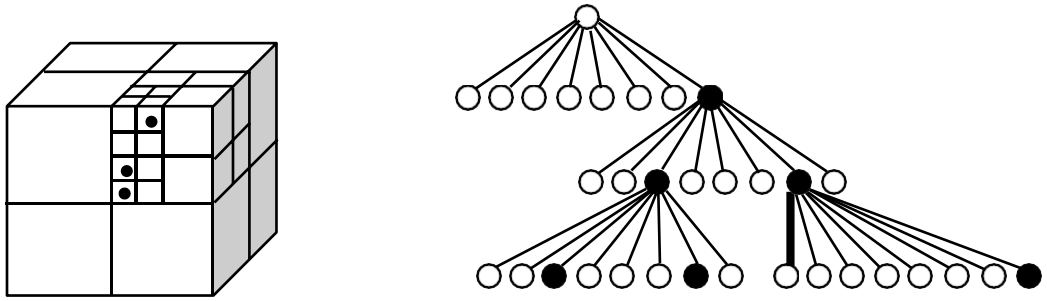


**Figure 4.1:** Octree structure and corresponding voxels in 3D space.

Regarding the color attribute, the average color of all the points belonging to a leaf voxel is computed and coded. In this work, since only geometry quality metrics and codecs are considered, this average color value is not encoded.

### 4.2.2 Graph-based Geometry Coding

It is well-known that the human visual system cannot perceive many high frequency details of a visual signal. This fact is used to design transform-based coding schemes, e.g., the point cloud geometry can be efficiently encoded by discarding some high frequency information. The main idea here is to assume that point coordinates in the 3D space are signals on a graph and obtain a frequency domain representation of the signal (3D locations) using a graph-based transform. This coding solution considers several steps:

1. **Point cloud clustering:** A clustering algorithm such as *K-means* is used to divide the reference point clouds into K clusters based on the geometry data. The main objective is that the size of the transform is small to be computationally tractable and also local as when a DCT transform is applied to fixed size blocks in 2D images.

2. **Graph creation:** Then, the graph structure is created, i.e., the edges of the graph are defined by connecting nearby points in each cluster. For each of the edges, a weight is assigned expressing the similarity between the corresponding signal values, in this case the 3D point coordinates according to a distance metric [90]. Thus, each point is connected to its *K* nearest neighbors and then any duplicated edges are pruned. When a sufficient number of neighboring connections is allowed, all points of a cluster belong to a single graph and, therefore, it is possible to obtain an efficient frequency representation.

3. **Graph transform** [90]**:** After obtaining a graph-based representation with the edges, vertices and associated weights, it is possible to compute the graph transform coefficients. The graph transform starts by computing an adjacency matrix *A*, which is populated with non-zero weights for every pair of connected points, i.e., for the pairs of points adjacent in the graph. Then, a graph Laplacian matrix *Q* is computed based on the weights and the eigenvector matrix of *Q* is applied as a transform to the 3D geometry coordinates. This transform is applied independently for the *x*, *y*, and *z* coordinates. Since the graph transform is based on the adjacency matrix *A*, the output is a set of DC and AC coefficients. When all the points in the cluster are connected directly or indirectly through another point, a single DC coefficient is obtained.

4. **Uniform quantization:** Uniform quantization with dead-zone is applied to the graph transform coefficients obtained from the graph transform applied to each cluster. The dead-zone is twice as large as the step size selected by the user. This quantization is responsible for the loss of quality and the appearance of compression artifacts at the decoder side, but also for the reduced rate.

Naturally, at the decoder, it is necessary to perform the inverse set of operations with respect to the encoder. Therefore, inverse quantization is performed as well as inverse graph transform. In the octree coding solution, the decoded point cloud does not have the same number of points as the reference point and, thus, spacing between points may be visible after rendering; moreover, the rendering primitive used (point, cube, or sphere) may also be visible, especially at low bitrates. On the other hand, for graph-based coding solutions, the number of decoded points is always the same as in the reference point cloud; however, due to the clustering process, some blocking artifacts may appear in the rendered point cloud as well as some holes between clusters.

## 4.3 Subjective Assessment Framework

In this section, the subjective test setup, and the creation of the video content for the subjective assessment are described.

### 4.3.1 From Point Clouds to Video Sequences

Three point clouds from the inanimate object dataset, notably *Statue Klimt*, *Shiva* and *Egyptian Mask*, and three point clouds from people object dataset, notably *Loot*, *Long dress* and *Red and Black*, were selected from the MPEG repository [91]. These point clouds are shown in Figure 4.2. These two datasets have rather different characteristics due to the way its acquisition was made. While the first dataset has some missing points (holes) in localized regions and acquisition noise, the second dataset is cleaner without visible artefacts. Noisy point clouds may impact the subjects' quality opinion, and this applies for both the reference and decoded point clouds, thus making the reference already annoying in terms of quality. All point clouds are static and have been coded using the two selected geometry coding solutions to obtain decoded point clouds with low, medium, and high quality. After, each point cloud was rendered to obtain a format suitable for visualization by the final user. In this case, both the reference and decoded point clouds were processed by the Technicolor renderer, currently known as Interdigital renderer [92].

**Figure 4.2:** PCs selected for subjective assessment. From left to right: *Loot*, *Red and Black*, *Long Dress*, *Shiva*, *Statue Klimt*, and *Egyptian Mask*.

To render a point cloud into a video sequence it was first necessary to define a virtual camera path for the user visualization of the object. Considering that the selected point clouds correspond to individual objects, it was considered appropriate to define a spiral camera path around the object, starting from a full object view and going to a closer look, with 512 views, i.e., around 17 seconds. Figure 4.3 shows the camera positioning from two different angles.

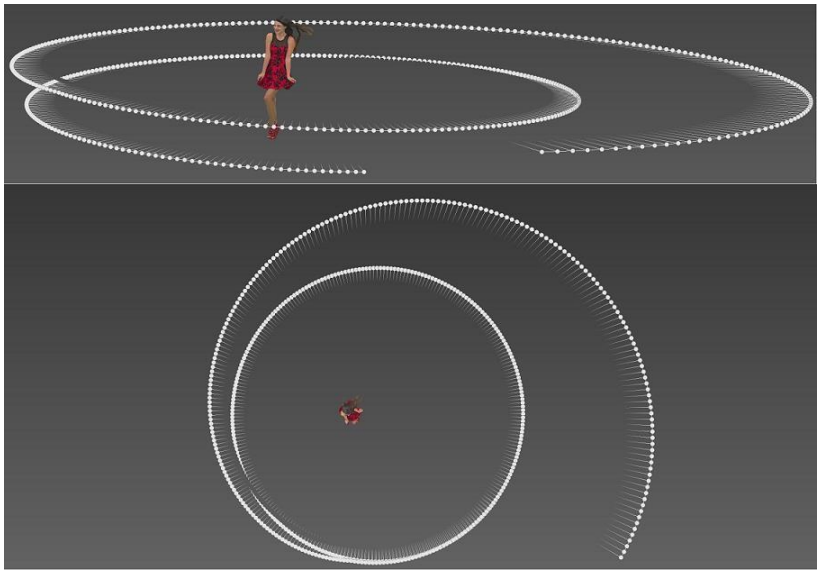


**Figure 4.3:** Camera path used for the point cloud rendering process.

A key parameter of the point cloud rendering process is the type of rendering primitive for each point; the most used primitives are the cube, point and splats. After some experiments, the cube primitive was selected. Cubes are similar to pixels in images and create better quality images by better filing the area between points. The size of the cube primitive for each point is also an important parameter; while using a smaller size may create holes, thus making the object surface less smooth, using larger sizes may lead to lack of detail. Therefore, the following method was adopted to determine the point size for the octree-based codec:

1. Each point cloud was normalized and scaled to its bounding box.
2. The mean distance of each point to its 10 nearest neighbors was computed.
3. The average distance over the full point cloud was selected as the point size for rendering.

This method does not work well for point clouds encoded with the graph transform-based codec because of the larger distances around the clusters' boundaries. The average nearest neighbor distance is a good estimation of point distance in a point cloud. However, this distance cannot be used since the distance among clusters which are created due to graph transform coding is rather different (usually it is much larger), i.e., it does not follow the same distribution. Therefore, for the graph transform-based codec, the point size obtained from the above process (i.e., for octree coding) needs to be mapped to a suitable point size for graph-based coding. In this case, it was found that using the following curve expressed by the division of two polynomials leads to good results:

$$f(x) = \frac{ax^2 + bx + c}{x + d} \tag{4.1}$$

The parameters *a*, *b*, *c*, and *d* were found experimentally using a set of point sizes considered optimal in a subjective way in a short, small informal subjective test, thus obtaining the values: *a*=3.7, *b*=6.7, *c*=3.8 and *d*=-0.8. After setting the primitive and point size, point clouds and their associated colors (color was not coded) were rendered to a raw RGB video sequence with 800×600 spatial resolution at 30 fps, leading to a duration of 17 s using the predefined spiral camera path.

### 4.3.2 Subjective Assessment Methodology

Since the two selected point cloud datasets have rather distinct characteristics, the subjective assessment was organized in two separate sessions. For both sessions, the DSIS methodology was adopted to obtain the subjective scores. Subjects included 4 females and 18 males. The goal of the experiment was briefly explained to the subjects, and they were asked to participate in a short training session just before each one of the test sessions. For the training sessions, the *Head* and *Queen* point clouds were used [91]. The point cloud reference and impaired rendered videos were viewed one after the other and scored by each subject with the play of each new video sequence controlled by the subjects as they had to press 'Play' to see a new sequence. The duration of the whole test was around 30 minutes (≈ 8 minutes of train and ≈ 22 minutes of test sessions). All rendered point clouds were organized into 9 scoring rounds within each session; each round included all point clouds with a random level of quality. According to Recommendation BT-500.13 [31], the subjects visualized first the reference/original rendered point cloud and after the corresponding geometry impaired/decoded point cloud (always using the original color); the impairments were scored in a 1-5 scale associated to five quality levels, notably 'very annoying', 'annoying', 'slightly annoying', 'perceptible but not annoying' and 'imperceptible'. The *IST Point Cloud Subjective Assessment Framework* developed in [93] was used to perform the visualization for the test sessions. Figure 4.4 shows an example of one of the point clouds in the *IST Point Cloud Subjective Assessment Framework*.

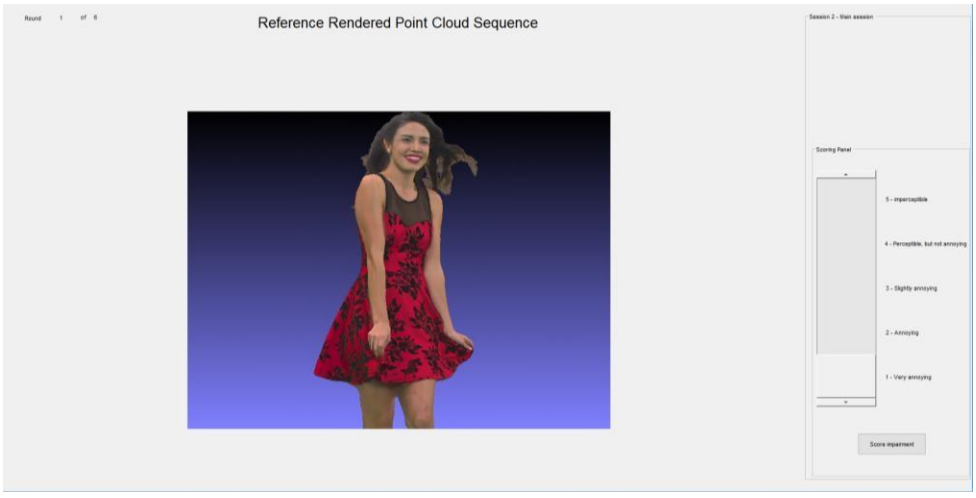



**Figure 4.4:** *IST Point Cloud Subjective Assessment* Framework: the scoring panel is shown only after viewing both the reference and decoded videos.

The rendered point clouds were visualized on a 23-inch ASUS VH238 monitor with 1920×1080 spatial resolution. An i7 workstation with the Intel HD 530 graphic card and 128MB video memory was used to play the test sequences at the correct frame rate.

**Subjective Scores Processing**

The same outlier removal process as described in Chapter 2 was deployed before further processing the subjective scores. In this study, one of the subjects was rejected. After the outlier removal procedure, the average of all scores across the subjects was calculated for each test sequence, for each coding rate; this corresponds to the MOS for each (sequence, rate) pair.

## 4.4 Performance Evaluation

This section focuses on the evaluation of the objective-subjective correlation of the selected point cloud geometry objective quality metrics for the selected point cloud geometry codecs, thus considering the artefacts present in the decoded point clouds. First, Section 4.4.1 presents the coding parameters used for the experiments; next, Section 4.4.2 presents the objective- subjective correlation scores for the most popular objective quality metrics at the time.

### 4.4.1 Coding Conditions

The reference and decoded point clouds used in the subjective quality assessment tests have different dynamic range and point densities. Therefore, the coding parameters had to be adjusted for each point cloud to create three clearly distinguishable qualities. For the graph transform-based codec, each object was divided into 1000 clusters. Table 4.1 shows the coding parameter values used to create the decoded point clouds for the subjective assessment experiment, notably the quantization parameter for the graph transform-based (GT) codec ($QP$) and the voxel size ($v_s$) for the octree-based codec. Since point clouds have different number of points, densities and other characteristics, these coding parameters where manually identified to obtain a wide range of qualities.

### 4.4.2 Objective Quality Metrics Evaluation

In this chapter, the same correlation performance metrics as in Chapter 3 are used to evaluate the correlation performance for the objective quality, notably PLCC, SROCC and RMSE. Since at the time of this work, point clouds had floating point precision (and were not voxelized), they were normalized before computing the RMSD and Hausdorff distances in order to compute the distortion on a similar scale for the set of point clouds under evaluation.

To have the same scale for the objective quality metrics and MOS values, a nonlinear logistic function was applied to the objective quality scores, as recommended in [32]. Let $Q_i$ be a specific objective quality score for video $i$ and $\beta_k$ the logistic function parameters. Based on [32], the predicted subjective score $MOS_p^i$ for the video sequence $i$ is computed from the quality score $Q_i$ of some objective quality metric according to:

$$MOS_p = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-\left(\frac{Q_i - \beta_3}{\beta_4}\right)}} \tag{4.3}$$

To estimate the $\beta_1$, $\beta_2$ and $\beta_3$ logistic function parameters, a non-linear regression of the objective quality metric scores over the MOS was performed. This allows fitting the objective quality metric scores to the MOS values. This fitting was performed using the *MATLAB* non-linear regression function, with initial parameter estimates as suggested in [32]. Figure 4.5 shows the predicted subjective scores (based on the objective scores) versus the real MOS values for all the decoded point clouds (with the two selected codecs) in this experiment. Since there were two test sessions, one for the people dataset and another for the inanimate dataset, with different characteristics, the two cases are treated, first independently, and later jointly.

**Table 4.1:** Point cloud geometry coding parameters.

| **Dataset** | **Point Cloud** | **No. Original Points** | **Quality** | $QP$ | $v_s$ |
|---|---|---|---|---|---|
| People | Loot | *~782,000* | High | 10 | 11 |
| | | | Med. | 30 | 21 |
| | | | Low | 60 | 35 |
| | Red and Black | *~700,000* | High | 2 | 13 |
| | | | Med. | 10 | 23 |
| | | | Low | 20 | 40 |
| | Long Dress | *~800,000* | High | 1 | 12 |
| | | | Med. | 10 | 19 |
| | | | Low | 30 | 35 |
| Inanimate objects | Statue Klimt | *499,886* | High | 0.05 | 0.05 |
| | | | Med. | 0.15 | 0.13 |
| | | | Low | 0.35 | 0.25 |
| | Egyptian Mask | *272,689* | High | 0.05 | 0.08 |
| | | | Med. | 0.15 | 0.14 |
| | | | Low | 0.25 | 0.4 |
| | Shiva | *1,010,591* | High | 0.01 | 0.04 |
| | | | Med. | 0.05 | 0.05 |
| | | | Low | 1 | 0.1 |

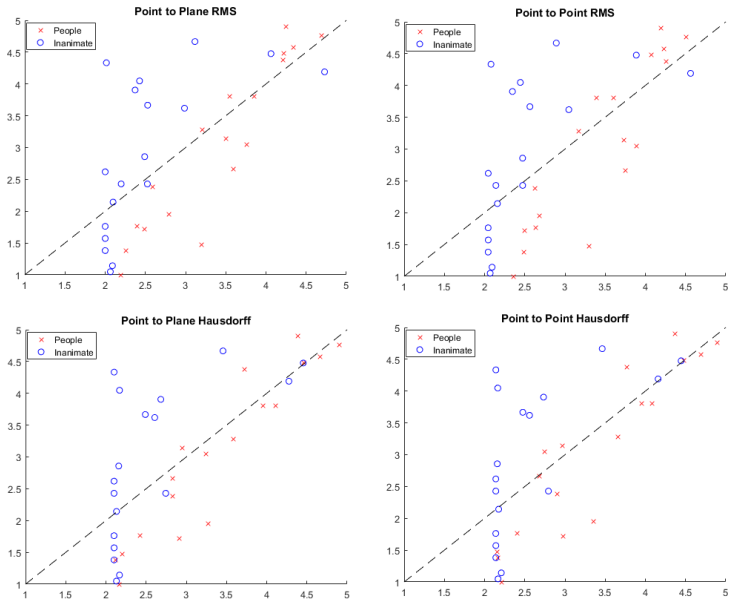



**Figure 4.5:** MOSp (predicted objective scores) vs. MOS (actual values).

**Table 4.2:** PLCC, SROCC and RMSE results for all objective quality metrics and point cloud datasets.

| | | | PLCC | | | SROCC | | | RMSE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Codec:** | | | **Octree** | **GT** | **ALL** | **Octree** | **GT** | **ALL** | **Octree** | **GT** | **ALL** |
| **Inanimate** | **RMSD** | ***Po2Po*** | 0.92 | 0.82 | **0.75** | 0.88 | 0.61 | **0.70** | 0.52 | 0.58 | **0.79** |
| | | ***Po2Pl*** | 0.91 | 0.82 | 0.74 | 0.88 | 0.77 | 0.67 | 0.56 | 0.58 | 0.81 |
| | **Hausdorff** | ***Po2Po*** | 0.89 | 0.90 | 0.65 | 0.85 | 0.80 | 0.52 | 0.60 | 0.44 | 0.91 |
| | | ***Po2Pl*** | 0.90 | 0.90 | 0.67 | 0.85 | 0.80 | 0.64 | 0.58 | 0.45 | 0.89 |
| | **Number of Points** | - | 0.78 | - | - | 0.73 | - | - | 0.83 | - | - |
| **People** | **RMSD** | ***Po2Po*** | 0.99 | 0.97 | 0.89 | 0.98 | 0.92 | 0.87 | 0.15 | 0.31 | 0.57 |
| | | ***Po2Pl*** | 0.98 | 0.97 | 0.93 | 0.98 | 0.93 | 0.93 | 0.22 | 0.33 | 0.46 |
| | **Hausdorff** | ***Po2Po*** | 0.98 | 0.99 | 0.92 | 0.94 | 0.93 | 0.91 | 0.24 | 0.18 | 0.50 |
| | | ***Po2Pl*** | 0.97 | 0.99 | **0.94** | 0.94 | 0.97 | **0.94** | 0.25 | 0.20 | **0.43** |
| | **Number of Points** | - | 0.91 | - | - | 0.82 | - | - | 0.45 | - | - |
| **All** | **RMSD** | ***Po2Po*** | 0.85 | 0.64 | 0.67 | 0.82 | 0.50 | 0.65 | 0.67 | 0.90 | 0.92 |
| | | ***Po2Pl*** | 0.85 | 0.64 | 0.71 | 0.82 | 0.57 | **0.70** | 0.67 | 0.90 | 0.87 |
| | **Hausdorff** | ***Po2Po*** | 0.81 | 0.67 | 0.72 | 0.79 | 0.58 | 0.66 | 0.74 | 0.87 | 0.85 |
| | | ***Po2Pl*** | 0.82 | 0.66 | **0.73** | 0.79 | 0.61 | **0.70** | 0.72 | 0.88 | **0.84** |
| | **Number of Points** | - | 0.80 | - | - | 0.80 | - | - | 0.76 | - | - |

Table 4.2 shows the objective-subjective correlation performance for the four selected objective quality metrics (the same as for Chapter 3 and described in Section 2.6.3) in terms of PLCC, SROCC and RMSE, for the two codecs and the two and full datasets; the best correlation performance values (for the ALL column) are highlighted in bold for the inanimate, people and all datasets. From the results, the following conclusions may be derived:

- **Graph-based transform vs. octree codecs:** Both the Po2Po and Po2Pl quality metrics perform rather well when fitting is performed, if the MOS values for each codec are independently considered. For the graph transform-based codec, the objective-subjective correlation is poorer than for the octree codec due to the blocking artifacts at the cluster borders. However, since ideally a quality metric must be agnostic to the type of artifacts introduced, the key results are the correlation values for the two codecs together (ALL column); as shown, this correlation decreases significantly (up to 0.25 points in *PLCC*) when compared to the separate codecs performance. This also means that there is still room to improve the point cloud objective quality metrics.
- **Po2Po vs Po2Pl metrics:** The objective-subjective correlation results show that the Po2Po and Po2Pl quality metrics have similar performance for the same distance metric. Note that while the objective quality scores for these metrics are different, their correlation with the MOS values is similar. This also means that, for the distortions caused by the types of coding solutions tested, modelling the point cloud surface as a plane and measuring the distance to this plane does not bring many improvements; this is rather true when the nearest point is in the direction of the normal of the vector, as it occurs for some coding artifacts. This may change if other codecs and coding artifacts are considered as in the future MPEG G-PCC and V-PCC codecs.
- **RMSD vs. Hausdorff distances:** The difference between the RMSD and Hausdorff distances correlation performances is not significant for the people dataset, which includes rather dense and regular point clouds. However, for the inanimate dataset, the difference is more significant since the holes present may influence the maximum distance and may lower the correlation of the Hausdorff-based metric with the MOS.
- **Dataset influence:** Point clouds in the people dataset have higher perceptual quality after rendering, due to their higher density and lack of acquisition noise, in comparison with the inanimate dataset (even for the reference point clous). Therefore, higher objective-subjective correlation is obtained for the people dataset when all codecs are considered. This is mainly due to the fact that the noisy surfaces with many holes after rendering of the inanimate dataset objects may make the scoring more challenging for the subjects. Note that the objective quality scores do not consider the intrinsic quality of the reference point clouds and 3D rendering since a DSIS methodology has been adopted.

As a curiosity, Table 4.2 also assesses the correlation performance of the number of decoded points if it is taken as an objective quality metric. This could be a rather good metric to assess the quality of decoded point clouds with the octree codec for which quality is dependent on the number of

points. However, it only works well in this specific case because the rate and, therefore, the quality in octree-based coding are directly associated with point reduction, i.e., lower rates lead to lower number of points. For the graph-based codec (and all codecs for which this relation does not apply), this metric would not be useful since the number of decoded points is always the same for this specific codec.

## 4.5 Final Remarks

This chapter targets the objective-subjective correlation evaluation of the geometry point cloud quality metrics available at the time this study was performed, namely Po2Po and Po2Pl RMSD and Hausdorff distances. To assess the correlation performance of these quality metrics, two point cloud codecs were selected based on two different coding approaches: the organization of the points in an octree structure and the spectral representation of a graph connecting nearby points. Experimental results have shown that the correlation between objective and subjective scores is similar, although with a slight advantage to the Po2Pl metrics when the two codecs and different datasets are considered. RMSD as a metric is slightly better than Hausdorff distance when all data and both codecs are considered.

The work in this chapter let to a conference publication, notably:

**A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Subjective and objective quality evaluation of compressed point clouds," *IEEE Workshop on Multimedia Signal Processing*, Luton, UK, October 2017.

# Chapter 5

# Impact of Rendering on Perceived Quality of Point Clouds

## 5.1 Context and Objectives

Point clouds can be visualized on a variety of devices, such as 2D displays, head-mounted displays (HMDs), augmented reality devices and even on stereoscopic or multi-stereoscopic displays. However, independently of the type of display, point clouds cannot be directly visualized and require a rendering technique to create the data that may be visualized; this may be seen as a post-processing step after decoding. Nowadays, there are several point cloud rendering approaches [94] [95] that may significantly impact the perceived point cloud quality in different ways. Often, the subjective and objective quality assessment studies available in the literature do not use the same type of coding and rendering solutions as well as test conditions and thus, reach different conclusions and are difficult to compare. Therefore, it is critical to study the impact of different rendering approaches on the subjective and objective quality of decoded point clouds under meaningful and representative test conditions.

On the other hand, many relevant works on subjective and objective quality assessment [96] - [97], [98], [99] and [100] rely on rather simple coding solutions such as *octree pruning*, which are inefficient compared to the state-of-the-art and produce a rather distinctive type of artefacts. However, more sophisticated, and more efficient lossy point cloud geometry coding solutions (some already standardized) are now available, which produce decoded point clouds with very different characteristics and artefacts. As an example, some point cloud geometry codecs significantly increase the number of decoded points to hide coding artefacts, thus achieving a better perceived quality. This makes the subjective and objective quality assessment of point clouds more complex, especially when more efficient coding and rendering solutions are considered.

While point clouds have commonly two major components, geometry, and color (or texture), this work focuses on the perceived quality impacts of degradations on the geometry component of the point cloud representation. Geometry artefacts are very important for the final perceived quality since this type of degradations may reduce the realism of the decoded geometry, e.g., due to the appearance of holes, and deformed and noisy surfaces, consequently leading to poorer user immersion. Figure 5.1 shows an example of geometry artefacts, in this case associated to the MPEG G-PCC codec (original texture was used for recoloring), which results in a rather low perceived quality, very much depending on the color masking effect. Despite its impact on the perceived quality, geometry degradations have not been much addressed in the literature. Besides, while the geometry is an essential component of the point cloud representation, the color attributes (which are optional) may not be available due to limitations in the acquisition process, e.g., point clouds acquired by LIDAR only devices. However, as will be seen later in this chapter, when present, color has an important masking effect regarding the geometry artefacts.

**Figure 5.1:** *Arco Valentino* point cloud: left) reference point cloud; and right) MPEG G-PCC decoded point cloud with lowest rate. Reference texture used for recoloring.

In this context, the main objectives of this work are:

- **Point cloud rendering after coding – subjective quality assessment:** Study of the subjective quality impact of multiple coding and rendering combinations for relevant, lossy point cloud coding and rendering solutions. Moreover, the visibility of the distortions associated to each point cloud geometry codec under different rendering scenarios will be analysed. This first goal is critical for the design of a suitable point cloud visualization solution as a practical, effective rendering solution must be chosen.
- **Point cloud rendering after coding – objective quality metrics assessment:** Evaluation of the correlation performance of available point cloud objective quality metrics for multiple coding and rendering combinations, i.e., for different types of rendering and coding artefacts. This should allow understanding the strengths and weaknesses of available objective quality metrics as well as their scope of validity, i.e., for which conditions these metrics represent well enough the human perceived quality. This second goal is critical for the design of more reliable

point cloud objective quality metrics, notably for the evaluation of new point cloud geometry coding solutions as well as associated techniques.

- **Rendered point cloud quality assessment dataset**: Provision of the first public dataset of mean opinion scores (MOS) and corresponding decoded point clouds using relevant, lossy point cloud geometry coding solutions. These point cloud geometry codecs produce a distinctive set of artefacts that were not considered when the most popular point cloud geometry objective quality metrics were designed. This third goal is particularly important for the point cloud quality assessment community, since not many subjective quality studies are available at the time this work was performed and, from those available, none allowed to assess the impact of the rendering process that is always performed after decoding.

This was the first work where the coding and rendering processes, which play a major role on the final perceived quality, were jointly evaluated, in this case for static point clouds and geometry coding. To be able to isolate and, thus, directly assess the impact of the geometry artefacts on the final perceived quality, no color attributes coding was considered in this work although the original color may be used for rendering after recoloring.

This chapter is organized as follows. Section 5.2 describes the point cloud rendering solutions used in the subjective experiments of this chapter. Section 5.3 introduces the codec used in the experiments of this chapter and describes the artefacts associated with three selected point cloud geometry coding solutions and illustrates them when three point cloud rendering solutions are used. Section 5.4 presents the subjective test framework that is used for further studies in this chapter. Section 5.5 describes the subjective quality evaluation study along with some key conclusions and Section 5.6 describes to the correlation evaluation for the most relevant point cloud objective quality metrics. Finally, Section 5.7 reports the main conclusions and final remarks.

## 5.2 Point Cloud Rendering Solutions

In this chapter, the impact of rendering on the perceived quality of a point cloud is studied. In this context, for the experiments of this Chapter it was used one solution for each of the three types of rendering approaches explained in Section 2.4, namely, *RPoint*, *RColor* and *RMesh*. In the following, important information about the representative solution (not previously described) for each of these rendering approaches, are presented.

### 5.2.1 *RPoint* Rendering

For this point cloud rendering solution, the selected rendering primitive was a square because they are the smallest element of a 2D image (pixels); the point size was set to the minimum value able to fill the 3D space between points completely, thus avoiding holes, leading to a so-called watertight rendering.

Regarding shading, color attributes are not used in this specific rendering approach, in order the impact of geometry distortions may be assessed without any additional component if it exists. To obtain the normal vectors, a (best fitting) plane was used as the local surface model and an automatic estimation for the neighborhood radius was used, as suggested in [14]. This automatic estimation helps to find a suitable radius as a too small radius may result in some points having an invalid normal and a too large radius may result into smoothed edges. By fitting a local surface, only the direction of the normal can be computed and, thus, the orientation of the normal was determined with the minimum spanning tree algorithm [15].

### 5.2.2 *RColor* Rendering

For this point cloud rendering solution, to isolate the impact of geometric distortions, the color attributes are not coded and, thus, the original color is used to recolor the decoded point cloud geometry. The recoloring process occurs when the number of points in the decoded point cloud is different (or the same) from the original number of points. The recoloring procedure uses the original color and performs a mapping of the original colors in the original positions to the decoded points positions. In this case, the vertex attribute transfer method available in *MeshLab* [101] was used for the recoloring process. Since the color information already consists of lighting information from the acquisition setup, we did not shade points in this approach to preserve as much as possible the color fidelity of the point cloud representation.

### 5.2.3 *RMesh* Rendering

For this point cloud rendering solution, the procedure to reconstruct the surface proposed in [97] was followed. The Poisson Surface Reconstruction algorithm, available in the popular CloudCompare [83] software, was selected with default parameters. The estimation of the normal vectors was performed as for the RPoint solution; no color attributes were used to be able to directly assess the subjective impact of the geometric artefacts.

## 5.3 Point Cloud Coding Artefacts after Rendering

This section describes the artefacts associated with each of the selected point cloud geometry coding solutions after rendering since this is what the users see. Three point cloud geometry coding solutions were selected for the studies performed in this chapter, namely the octree-based coding solution from the PCL Library [89], the MPEG G-PCC standard [102] and the Intra mode of the MPEG V-PCC standard [103]. A detailed explanation of these codecs can be found in Section 2.3. A characterization of the coding artefacts is important to understand the perceptual impact in the subjective tests and the limitations of the available objective quality metrics. With this purpose in mind, some frames were extracted from the videos rendered for the subjective test sessions described in Section 5.4.2. To study the impact of rendering, the videos created for the subjective test sessions were obtained with the three rendering approaches presented in Section 2.4 labelled as *RPoint*, *RColor*, and *RMesh*. To better assess the coding artefacts, the geometry of all point

clouds was coded using the lowest rate test conditions as described in Section 5.4.1. The selected point cloud examples try to show as much as possible the most impacting visual artefacts found during this study.

### 5.3.1 PCL Codec Artefacts

For the PCL codec, as the target bitrate decreases (leading to a lower octree depth), the number of decoded points also decreases since all points inside a voxel are represented by just one point at the voxel center. The consequence is an increase of the distance between decoded points and thus lack of detail. When PCL decoded point clouds are rendered, for any rendering solution, the lack of detail (i.e., points) results into a pixelated (or overly sub-sampled) decoded point cloud. An example of the artefacts produced with PCL coding at low rate is illustrated in Figure 5.2, for three point clouds and all rendering solutions. As shown, the rendered point clouds are rather pixelated (*RPoint* and *RColor*) or lack detail (*RMesh*).

### 5.3.2 MPEG G-PCC Codec Artefacts

The MPEG G-PCC codec prunes the octree at some specific depth and creates after a so-called *Trisoup* surface, representing the points at that depth with more precision. The rendering artefacts produced are very different from the PCL codec, since the number of decoded points is no longer reduced. An example of the artefacts produced by G-PCC at low rate is illustrated in Figure 5.3 for three point clouds and all rendering solutions. The geometry artefacts essentially come from the *TriSoup* process which may create false edges at the block boundaries or triangles; for low rates, these triangles may be very visible even with texture masking. Moreover, when the point cloud is sparse in some region, the *TriSoup* process may cause artificial holes (with polygonal shapes) or even an increase of the size of the holes already present in the reference point clouds.

### 5.3.3 MPEG V-PCC Codec Artefacts

For MPEG V-PCC in Intra mode, point cloud data is first 2D mapped and after coded with traditional Intra prediction and 2D transform tools. The more visible rendering artefacts correspond to blockiness and false edges, often associated to the directional Intra prediction modes. An example of the rendering artefacts produced by V-PCC is illustrated in Figure 5.4 for three point clouds and all rendering solutions. While false edges are visible, mostly for *RPoint* rendering, V-PCC distortions are not very visible for *RColor* rendering, notably due to the texture masking effect. For *RMesh*, the entire decoded point cloud is smoother compared to *RPoint*. However, some details are lost, which may cause lower perceived quality.


**Figure 5.2:** PCL coding artefacts: a) *Loot*; b) *Egyptian Mask*: and c) *House without a Roof*, when rendering with *RPoint*, *RColor* and *RMesh*.


**Figure 5.3:** G-PCC coding artefacts: a) *Loot*; b) *Egyptian Mask*; and c) *House without a Roof*, when rendering with *RPoint*, *RColor* and *RMesh*.


**Figure 5.4:** V-PCC coding artefacts: a) *Loot*; b) *Egyptian Mask*; and c) *House without a Roof*, when rendering with *RPoint*, *RColor* and *RMesh*.

## 5.4 Subjective Assessment Framework

In this section, the creation of the visual content for the subjective experiments is described along with the test setup and conditions, in summary the subjective assessment framework.

### 5.4.1 Test Conditions

Six static point clouds have been selected from the MPEG content repository [104] , notably *Egyptian Mask* and *Frog,* from the inanimate objects class, *Facade9* and *House without a Roof*, from the buildings and facades class, and *Longdress* and *Loot,* from the people class. This selection includes point clouds with different levels of complexity (as defined by MPEG in [104]), notably with four point clouds from class A (lowest complexity) and two point clouds from class B (highest complexity for static point clouds). These six selected point clouds have rather different characteristics in terms of content, geometry, and color. The most important factors for the point clouds selection process were: i) the point cloud density, i.e., sparse, and dense point clouds; ii) the (semantic) type of content, i.e. (point clouds from inanimate objects, facades and buildings and people classes; iii) the point cloud geometry structure, i.e., point clouds with holes and with flat surfaces; and iv) the color characteristics, i.e., (point clouds with small or large color gamut. Table 5.1 shows the point cloud characteristics, notably name, number of points, coordinates precision and category while Figure 5.5 shows *RColor* rendered views from the reference point clouds.

**Table 5.1:** Test point clouds and associated characteristics.

| Point Cloud Name | No. Points | Precision | Category |
|---|---|---|---|
| *Egyptian Mask* | 272,684 | 12-bit | Inanimate objects |
| *Facade9* | 1,596,085 | 12-bit | Facades and buildings |
| *Frog* | 3,614,251 | 12-bit | Inanimate objects |
| *House without a roof* | 4,848,745 | 12-bit | Facades and buildings |
| *Loot* | 805,285 | 10-bit | People |
| *Longdress* | 857,966 | 10-bit | People |

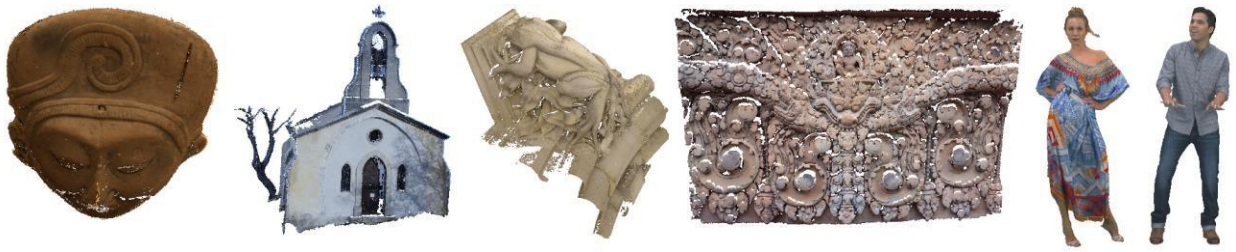


**Figure 5.5:** Test materials with *RColor* rendering. From left to right: *Egyptian Mask*, *House without a Roof*, *Frog*, *Facade9*, *Longdress*, and *Loot*.

The selected point clouds were coded with the three selected point cloud codecs, at three different rates, to obtain decoded point clouds with three different perceptual qualities, labelled as Low (L), Medium (M) and High (H). The selected codecs represent three different coding approaches, notably PCL for octree structures, MPEG G-PCC for surface models and MPEG V-PCC for projection-based coding. For each of the MPEG point cloud geometry codecs, three different rate points have been selected following the suggested coding parameters in the MPEG Common Test Conditions (CTC) [104] for lossy coding. These rate points resulted into three distinguishable qualities, ranging from low to high. For PCL, the octree depth parameter was controlled in a way to obtain a similar range of qualities compared to V-PCC. Table 5.2 shows the coding parameters used for the PCL and MPEG G-PCC codecs. For G-PCC, the octree depth establishes the point cloud precision (after the voxelization step). The *level* parameter corresponds to some octree layer after which the Trisoup polygonal/surface representation is used. For PCL, the octree depth (*OD)* is set indirectly, using the PCL Octree Resolution (*OR*) parameter, which corresponds to the size of the voxel computed as $OR = 2^{(P-OD)}$, with *P* corresponding to the point cloud precision (defined in Table 5.1).

**Table 5.2:** Octree depth and level for PCL and G-PCC, for three different rates: Low, Medium, and High.

| Point Cloud Name | PCL | | | G-PCC | | | |
|---|---|---|---|---|---|---|---|
| | Octree Depth | | | Octree Depth | Octree Level | | |
| | L | M | H | L/M/H | L | M | H |
| *Egyptian Mask* | 7 | 8 | 9 | 9 | 5 | 6 | 7 |
| *Frog* | 8 | 9 | 10 | 11 | 7 | 8 | 9 |
| *House without a Roof* | 8 | 9 | 10 | 11 | 7 | 8 | 9 |
| *Facade9* | 8 | 9 | 10 | 11 | 7 | 8 | 9 |
| *Loot* | 7 | 8 | 9 | 10 | 6 | 7 | 8 |
| *Longdress* | 7 | 8 | 9 | 10 | 6 | 7 | 8 |

Table 5.3 shows the MPEG V-PCC HEVC quantization parameter (*QP)* used for depth map coding (note that no color coding was performed) and *B0* is the occupancy map precision. For V-PCC, all the test materials were voxelized to 10-bit precision.

**Table 5.3:** Quantization Parameter (QP) and Occupancy Map Precision (B0) for V-PCC codec for Low, Medium and High Qualities.

| Quality | Low | Medium | High |
|---|---|---|---|
| **QP** | 32 | 24 | 16 |
| **B0** | 4 | 4 | 4 |

### 5.4.2 Test Sessions

The subjective quality assessment was performed in three test sessions, each corresponding to a different point cloud rendering approach. Following Section 5.2 from Chapter 2, the test sessions have been labelled as:

1. ***RPoint* session:** point clouds rendered with point-based rendering with point shading and no color attributes.
2. ***RColor* session:** point clouds rendered with point-based rendering with the original color (after recoloring) and no shading.
3. ***RMesh* session:** point clouds rendered with mesh-based rendering with surface shading and no color attributes.

**From Point Clouds to Video Sequences**

The point clouds were visualized in a non-interactive way, which means the reference and decoded point clouds were rendered to standard video sequences and shown on a 2D display, thus implying no user interaction. The advantage of such approach is that all subjects in the subjective test see the same parts of the decoded point clouds, exactly in the same way, thus obtaining more reliable subjective assessment scores, assuming the rendering path is appropriate. The *CloudCompare* point cloud processing software was used for rendering with the point size, normal estimation and surface reconstruction performed as described in Section 5.2. The lighting conditions, which influence the shading process in *RPoint* and *RMesh,* correspond to the default conditions, this means ambient light source (sun light) and no spotlight. A simple camera path rotation around the object was used to create the 2D rendered videos; this path was found to allow a rather complete visualization of the point clouds and, most importantly, the coding artefacts under evaluation. For some point clouds (e.g., *Facade9*), no geometry was acquired for the back side and, thus, the rotation path was restricted to the frontal part of the object. The virtual distance between the point cloud and the camera did not change, similarly to standard image and video subjective test methodologies where the distance between the subject and the display is typically fixed.

The rendered videos have a spatial resolution of 1600×800, a temporal resolution of 25 frames per second (fps). Based on the number of stimuli and number of sessions, the duration of rendered videos is 10 seconds. Total duration of each session was almost 35 minutes. For all the three test sessions, the rendered videos were visualized on a 23-inch ASUS VH238 monitor with 1920×1080 resolution. An i7 workstation with the Intel HD 530 graphic card and 128MB video memory was used to play the rendered videos at the correct frame rate.

### 5.4.3 Subjective Quality Assessment Methodology

The point clouds selected for the subjective study have rather different characteristics. Due to the acquisition process, some reference point clouds may be rather noisy, e.g., MPEG cultural heritage and buildings sub-category may have holes, outliers or even positioning errors. Also, the density

(number of points per unit of volume) of the reference point cloud may have a significant impact on the perceived quality of the reference rendered point cloud. These two factors may affect the subjective scores given by the subjects. Since these issues affect both the reference and decoded point clouds, the DSIS subjective test methodology was selected for all the test sessions of this subjective study. Thus, subjects visualize first the reference and after the decoded rendered point clouds and score the impairment of the decoded point cloud relatively to the reference point cloud, which allows to mitigate the impact of acquisition artefacts and other reference point cloud characteristics since they are already present in the reference point cloud.

There were 20 subjects participating in each test session with 18 people participating in all the three sessions and four people in one or two sessions. At the beginning of each session, the goal of the subjective assessment experiment was explained to the subjects, and they were asked to participate in a short training session to become familiar with the application interface. For the training sessions, the *Statue Klimt* point cloud from the same MPEG repository was used.

The full set of rendered point cloud videos was organized into six rounds per session with each round including all point clouds with one of the three levels of quality. Since there were six point clouds coded with three different codecs for three rate points, $6 \times 3 \times 3 = 54$ stimuli were assessed in each session. According to Recommendation BT-500.13 [31], the subjects saw first the reference rendered point clouds and after the impaired (this means decoded) rendered point cloud and scored the later in a 1-5 scale associated to five impairment levels, notably ‘very annoying’, ‘annoying’, ‘slightly annoying’, ‘perceptible but not annoying’ and ‘imperceptible’. The display of each new rendered video was controlled by the subjects by pressing ‘Play’. The subjects had the option to replay both video sequences (reference followed by impaired point clouds) before giving the subjective score. This option allowed to reduce the cognitive load of the subjects and, thus, obtaining more reliable scores. Each session had a duration of approximately 28 minutes, considering the training and scoring times. To avoid that the results of one session influenced the results of another session, a minimum of 48h between test sessions was respected.

For each session, outlier subjects were identified based on the collected scores, following the procedure in Recommendation BT.500-13 [31]; only one outlier was identified in the *RMesh* session. After, the average of all scores across the subjects were computed for each test point cloud, thus obtaining a MOS for each point cloud under evaluation. The subjective scores for the three test sessions along with the reference and decoded rendered point clouds are publicly available at [105] and, thus, may be used by the research community.

## 5.5 Subjective Quality Assessment

The focus of this section is on the study of the impact of different point cloud rendering solutions on the user perceived quality for decoded point clouds with different coding artefacts. The obtained subjective scores are analyzed to assess the visibility of the different coding artefacts. The subjective scores obtained for the three test sessions will be the basis for this study; in this case,

the MOS values represent the similarity between the reference and decoded point clouds and not the intrinsic point cloud quality for which many other factors play a role.

### 5.5.1 Impact of Rendering on Perceived Point Cloud Quality

This section studies the impact of the three rendering solutions on the perceived point cloud quality. Note that, within each session, the rendering methods were not mixed and, thus, the subjects evaluated decoded point cloud videos for each rendering solution independently.

Figure 5.6 shows a chart with the 54 MOS for all point clouds within each test session (each associated to a rendering solution). In Figure 5.6**,** the MOS are sorted in ascending order, i.e. from lower to higher scores, each score is labelled with a rendered point cloud index and corresponds to a coding condition, i.e. a combination of a point cloud with a codec and a rate. To identify which are the most frequent MOS per session (data not shown in Figure 5.6**)**, Figure 5.7 shows the MOS distributions (number of votes for each score) given by the subjects in the three rendering sessions.

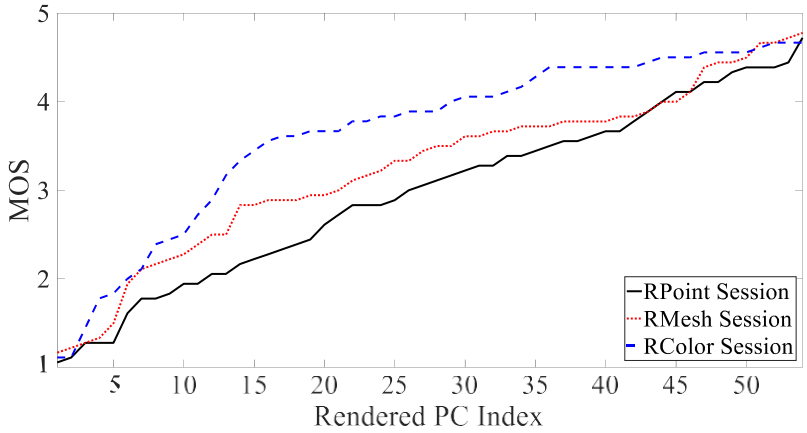



**Figure 5.6:** Sorted MOS for all test point clouds for the three test/rendering sessions.

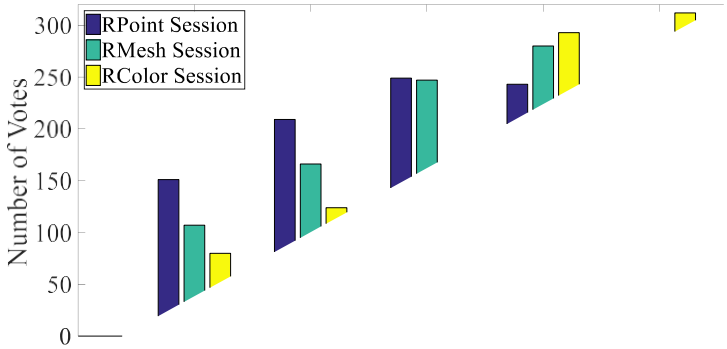



**Figure 5.7:** MOS histograms for the three test/rendering sessions.

Figure 5.6 shows that the scores are well distributed over the full range, from low (close to 1) to high (close to 5). The *RColor* session (blue curve) shows the highest MOS, followed by the *RMesh* session and, finally, the *RPoint* session. Based on the results, the following conclusions about impact of rendering, considering different rendering approaches, can be derived:

- ***RPoint* rendering:** The geometry coding distortions are more visible for *RPoint* rendering since *RMesh* and *RColor* have mechanisms to mitigate the visual impact of the coding artefacts, e.g., filtering or masking. This can be clearly observed in Figure 5.6 where the coding artefacts are more visible for the curve with lower MOS and, thus, as shown in Figure 5.7, more '1', '2' and '3' votes are obtained for *RPoint* compared to *RMesh* and *RColor*.
- ***RMesh* rendering:** As shown in Figure 5.7, *RMesh* rendering has higher MOS (and less low MOS) than *RPoint* rendering. This can be explained by the fact that *RMesh* rendering includes a surface reconstruction process (polygonal mesh creation) which smooths the point cloud and makes the coding distortions less visible, somehow behaving as a denoising filter. However, it should be emphasized that point cloud edges and details are also smoothed with this type of rendering and, thus, *RMesh* is not able to outperform a point-based rendering solution with color (*RColor*), where the points are simply rendered with a basic primitive. It also requires the extra processing step of surface reconstruction before rendering, which may be difficult to apply in some application scenarios due to the scene complexity or the point cloud size (number of points).
- ***RColor* rendering:** For *RColor* rendering, the original texture contains natural shading information, acquired from the light reflected by the object surface. This contrasts with the *RPoint* and *RMesh* renderings, which use a single color with synthetic shading and the final rendering result depends on the accuracy of the (extracted) normal vectors (geometry only). However, the original color captured during the acquisition can effectively mask many of the geometry distortions, changing the perceived surface of the objects. Also, the texture details can hide geometry distortions since the human visual system is less sensible to high frequencies; this will cause the subjects to perceive less distorted shapes and, thus, better scores are given in the *RColor* session. In fact, in this session, most of the scores are '4' and '5' as shown in Figure 5.7, which means that most of the decoded point clouds were considered to have low impairment regarding the reference point clouds. In this case, the most visible geometry distortions are limited to the object boundaries.

In summary, rendering with high quality color attributes masks the geometric distortions and results in higher perceived point cloud quality. However, color attributes may not be available, and some applications may require high geometry fidelity. For example, geographical information systems and cultural heritage applications typically only tolerate imperceptible geometry deformations; in such cases, *RPoint* rendering could be an appropriate choice to avoid the influence of color masking and geometry filtering. On the other hand, if color is not available and geometry degradations are tolerable if not visible, *RMesh* rendering should be used, since it allows to

mitigate the impact of some coding artefacts, e.g., holes and false edges, compared to *RPoint*, thus leading to higher perceived point cloud quality.

### 5.5.2 Impact of Rendering on the Coding Artefacts Visibility

This section studies the impact of the three rendering solutions on the visibility of the coding artefacts associated to the three selected codecs. With this purpose in mind, the MOS for the three point cloud codecs and the three rendering solutions are shown in Figure 5.8.

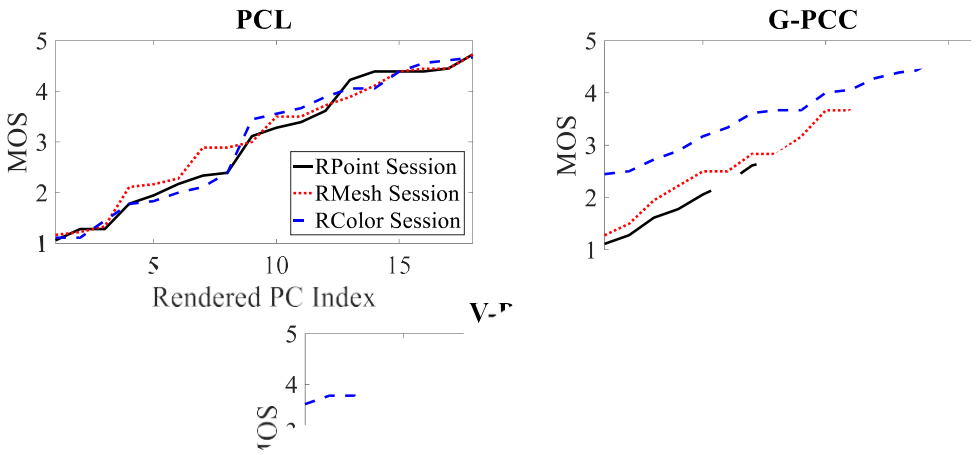



**Figure 5.8:** Sorted MOS for all point clouds for the three test/rendering sessions, separated for each codec.

This point cloud codec-based presentation of the MOS scores, thus more granular than Figure 5.6, allows comparing the impact of the rendering solution on the final perceived visibility when different coding artefacts are present. Figure 5.8 shows that the MOS distribution for each rendering approach is not similar for all codecs. The main conclusion is that the different types of coding artefacts are not equally visible for all rendering approaches. Based on the results, the following conclusions about the sensibility of the several rendering solutions to the point cloud codecs considered, and thus type of coding artefacts, may be derived:

- **PCL Coding:** PCL distortions are visible regardless of the rendering solution and, thus, MOS are rather well distributed in the 1-5 range. This is mainly because a pure octree point cloud coding solution controls the decoded quality by limiting its maximum depth and, thus, decoded point clouds have a lower number of points than the reference point clouds, sometimes significantly lower. Thus, larger point sizes for *RPoint* and *RColor* rendering are needed, thus creating a pixelated effect (perceptually unpleasant). Although a surface is reconstructed with

*RMesh* rendering, when the number of points is reduced, details are lost, and some meshes even show geometry distortions due to the surface reconstruction process.

- **G-PCC Coding:** G-PCC distortions are less visible for *RColor* rendering compared to *RPoint* and *RMesh*, since the color masks the surface distortions. However, false edges, holes and geometry distortions at boundaries are still visible for severe distortion cases. *RPoint* and *RMesh* follow a similar trend, with slightly better scores for *RMesh*, since it mitigates the impact of some coding artefacts (e.g., holes and false edges), thus offering a more visually appealing surface.
- **V-PCC Coding:** V-PCC distortions are not very visible for *RColor* rendering since they are not large enough to create strong deformations and the color masks most of the geometry distortions. Due to the V-PCC projection onto 2D maps (texture and depth) and the efficient HEVC coding process, most of the surfaces are consistently represented, although with some error regarding the original surface. V-PCC distortions are also less visible for *RMesh* than for *RPoint* rendering due to the impact of surface reconstruction-based rendering on the perceived quality.

### 5.5.3 Statistical Significance Analysis of Subjective Assessment Scores

This section presents a statistical significance analysis of the subjective quality assessment scores. The goal is to evaluate if the differences between the MOS for the three rendering approaches (*RPoint*, *RColor* and *RMesh*) are statistically significant at a given confidence level. Based on procedures suggested in previous works [106] [107] [108], three statistical tests were applied:

- **Welch ANOVA Significance Test** compares two means to see if they are equal. It is an alternative to the Classic ANOVA and can be used even if your data violates the assumption of homogeneity of variances (i.e., equal variances). If there are more than two groups of data, it can only show if the difference between group means is statically different or not. However, cannot identify which differences between group means are statistically significant. The null hypothesis in this test is that the groups of data are drawn from a distribution with the same mean.
- **Games-Howell Post-hoc Test** is applied after the ANOVA test and is used to compare all possible combinations of group differences when the assumption of homogeneity of variances is violated. However, it still requires the groups to follow a normal distribution. The null hypothesis in this test is that two groups of data in the pairwise comparison are drawn from a distribution with the same mean.
- **Wilcoxon Signed-Rank Test** is a non-parametric statistical hypothesis which is used to compare all possible combinations of group differences when normality of populations is violated in the groups of data. The null hypothesis in this test is that the groups of data are drawn from a non-parametric distribution with the same mean.

In null hypothesis significance testing, the p-value is the probability of the null hypothesis being correct. In all tests above, for α percent significance level, if the p-value is smaller than α, the null hypothesis is rejected.

To test if a group of data follows a normal distribution, Shapiro-Wilk statistical test is applied and to see if two (or more) groups of data have homogeneous variances, Levene's test is used in this section.

For all tests, the full set of obtained subjective scores was divided in three groups, one group of scores for each point cloud rendering approach, since the results being tested for statistical significance (sections 5.5.1 and 5.5.2) evaluate the impact of rendering. The selection of these statistical tests was motivated by the fact that for V-PCC data, the variance homogeneity test fails according to a Levene's test while the distribution of data is not normal for the 'All' case according to the Shapiro-Wilk normality test.

***Welch ANOVA significance test:*** To evaluate if the dependency of MOS values on the rendering method is statistically significant, the Welch ANOVA significance test was applied, thus comparing groups of MOS values, one group for each rendering method. This test measures the difference between the mean values of each group with a 5% significance level, without requiring homogeneity of variances. The null hypothesis assumes that MOS values for the various groups (rendering methods) are drawn from a population with equal means. Table 5.4 shows the p-values and associated MOS averages when considering all possible groups of scores for each codec ('PCL', 'G-PCC' and 'V-PCC' columns) and for all the codecs together ('All' column). When the p-value is lower than 0.05 (significance level), the separation between these rendering approaches is statistically significant.

**Table 5.4:** P-values for the Welch ANOVA significance test and MOS averages for each rendering session, i.e., *RPoint, RColor, RMesh*.

| | **PCL** | **G-PCC** | **V-PCC** | **All** |
|---|---|---|---|---|
| **p-value** | 0.98 | 0.019 | 0.0 | 0.002 |
| **Average MOS** | 3.0,3.0,3.1 | 2.8,3.8,3.2 | 3.1,4.2,3.5 | 3.0,3.7,3.2 |

***Games-Howell post-hoc and Wilcoxon signed-rank tests:*** To compare the several possible pairs of rendering methods, a multiple-comparison statistical test must be used. In this case, the Games-Howell post-hoc test was selected since it also does not require the homogeneity of variances, again with a 5% significance level. Table 5.5 shows the p-values obtained for this post-hoc test for all possible rendering pairs. Moreover, since the obtained MOS values do not follow a normal distribution (i.e., normality does not hold) for the 'ALL' case, the p-values obtained for the Wilcoxon signed-rank test (5% significance level) are shown in Table 5.6. The Wilcoxon signed-rank test assesses whether the group mean ranks differ. This test is more suitable for this case since

it is a non-parametric test, i.e., it does not assume any data distribution. For the Games-Howell post-hoc and Wilcoxon signed-rank tests, when the p-value is lower than 0.05 (significance level), there is statistical significance between groups of MOS. This means that MOS for those sessions can be considered different, regardless of MOS for each stimulus or even average of stimuli. e.g., MOS between V-PCC decoded point clouds is statistically different between *RPoint* and *RColor*, which in this case confirms that MOS in *RColor* is higher than *RPoint* (based on values).

**Table 5.5:** P-values for the Games-Howell post-hoc test for all rendering pairs (pair order is irrelevant).

| | **PCL** | **G-PCC** | **V-PCC** | **All** |
|---|---|---|---|---|
| ***RPoint↔RColor*** | 0.998 | 0.011 | 0.000 | 0.002 |
| ***RPoint↔RMesh*** | 0.973 | 0.586 | 0.144 | 0.300 |
| ***RColor↔RMesh*** | 0.988 | 0.162 | 0.000 | 0.085 |

**Table 5.6:** P-values for the Wilcoxon test for the 'All' case.

| | ***RPoint↔RColor*** | ***RPoint↔RMesh*** | ***RColor↔RMesh*** |
|---|---|---|---|
| **p-value** | 0.0 | 0.023 | 0.003 |

From the analysis of the results in Table 5.4 and Table 5.5, i.e., Welch ANOVA significance and Games-Howell post-hoc tests, respectively, the analysis of Section 5.5.2 can be confirmed and new conclusions may be derived:

- **PCL:** The difference between the MOS for the three rendering approaches is not statistically significant and thus, any rendering can be used. This was expected since PCL distortions are visible regardless of the rendering solutions and, thus, similar subjective scores were obtained for all rendering approaches.
- **G-PCC:** *RColor* is better than *RPoint* and, thus, if color is available, it should be used in point-based rendering solutions. There is no statistical difference between *RPoint* and *RMesh* and *RColor* and *RMesh*, meaning that there is no advantage in using mesh-based rendering (which may even require complex surface reconstruction methods).
- **V-PCC:** *RColor* is better than *RPoint* and *RMesh* and, thus, color effectively masks the geometric distortions associated to the V-PCC coding artefacts. For the 2$^{nd}$ best rendering method, there is no statistical difference between *RPoint* and *RMesh* and thus, this means that any of these two approaches may be used.

Finally, from the analysis of the results in Table 5.6 above, i.e., Wilcoxon signed-rank test results, statistical significance was obtained for all rendering pairs (i.e., *RPoint↔RColor*, *RPoint↔RMesh*, and *RColor↔RMesh*) for the 'All' case, meaning that a ranking order of the rendering methods is established. The results for this test show that *RColor* is statistically better

than *RMesh* and *RMesh* is statistically better than *RPoint*. This confirms the intuitive ordering shown in Figure 5.6 and the conclusions in Section 5.5.1.

## 5.6 Objective Quality Metrics Evaluation

In this section, the selected objective quality metrics performance will be presented and analyzed using the subjective scores obtained in the three test sessions, thus for different rendering approaches. As recommended in [31] [32], before assessing the objective quality metrics correlation performance, a nonlinear logistic fitting similar to Chapter 3 has been applied to the objective scores to map them to the subjective scores scale. To assess the objective quality metrics performance, the Pearson Linear Correlation Coefficient (PLCC) is computed as a measure of the linear dependence between the objective metric scores and the corresponding MOS.

Table 5.7 and Table 5.8 shows the PLCC and SROCC values for the 9 objective quality metrics described in detail in Chapter 2, section 2.6.3, for each rendering approach, independently computed for each point cloud codec and also considering all codecs simultaneously (column 'All'). With these correlation results, the performance of each metric can be assessed for each of the three test/rendering sessions described in Section 5.4.2. A detailed analysis of the results in Table 5.7 is presented in the following. First, from the perspective of the point cloud codec and coding distortions, next from the perspective of the rendering solution and, finally, assessing which objective quality metric performs the best and under which conditions.

**Table 5.7:** PLCC (%) between objective geometry quality metrics and MOS scores for the three selected rendering approaches. In bold, the best PLCC values and all the other PLCC values that do not deviate more than 0.02 from the best PLCC value.

| Type | Metric | ***RPoint* Session** | | | | ***RColor* Session** | | | | ***RMesh* Session** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **PCL** | **G-PCC** | **V-PCC** | **All** | **PCL** | **G-PCC** | **V-PCC** | **All** | **PCL** | **G-PCC** | **V-PCC** | **All** |
| **Point-to-Point** | **D1** | 84.5 | 53.7 | 26.3 | 51.9 | 84.1 | **85.5** | 44.1 | 64.3 | 90.5 | 32.7 | 7.5 | 39.0 |
| | **Hausdorff** | **90.5** | 45.6 | 34.2 | 23.7 | 87.1 | 59.2 | 57.8 | 18.6 | 88.3 | 49.4 | **31.2** | 32.5 |
| | **D1-PSNR** | 87.4 | **86.5** | **55.0** | 66.9 | 89.8 | 71.1 | 72.3 | **78.3** | **91.6** | 51.7 | 18.3 | 68.8 |
| **Point-to-Plane** | **D2** | 84.4 | 50.2 | 32.8 | 46.9 | 84.7 | 80.6 | 18.3 | 60.2 | 88.5 | 37.1 | 12.4 | 34.5 |
| | **Hausdorff** | 90.1 | 55.1 | 29.8 | 30.1 | 87.0 | 67.1 | 69.0 | 21.1 | 87.7 | 45.9 | 26.9 | 28.9 |
| | **D2-PSNR** | 90.1 | 82.4 | 52.1 | **69.7** | **90.3** | 54.6 | 61.4 | 78.2 | 91.0 | **63.2** | 28.0 | **72.1** |
| **Plane-to-Plane** | **MSE** | 72.4 | 55.5 | 54.1 | 51.8 | 55.7 | 68.3 | **74.5** | 24.7 | 40.0 | 28.0 | 28.7 | 30.3 |
| | **RMS** | 72.4 | 55.4 | 52.6 | 51.8 | 55.7 | 68.3 | **74.5** | 24.6 | 40.0 | 27.9 | 27.1 | 30.5 |
| | **MEAN** | 71.7 | 55.6 | 51.8 | 51.5 | 55.0 | 68.0 | 72.9 | 24.7 | 39.8 | 26.2 | 30.4 | 31.1 |
| - | **No. Points** | 65.2 | 21.8 | 27.9 | 12.3 | 69.0 | 27.1 | 60.6 | 43.1 | 68.6 | 33.1 | 28.9 | 2.3 |

**Table 5.8:** SROCC (%) between objective geometry quality metrics and MOS scores for the three selected rendering approaches. In bold, the best SROCC values and all the other SROCC values that do not deviate more than 0.02 from the best SROCC value.

| Type | Metric | *RPoint* Session | | | | *RColor* Session | | | | *RMesh* Session | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PCL | G-PCC | V-PCC | All | PCL | G-PCC | V-PCC | All | PCL | G-PCC | V-PCC | All |
| Point-to-Point | D1 | 73.1 | 65.9 | 32.7 | 61.8 | 76.2 | **91.0** | 59.1 | 74.2 | 84.5 | -1.5 | 2.5 | 43.3 |
| | Hausdorff | **84.4** | 36.3 | 37.7 | 25.2 | 79.6 | 60.8 | 55.5 | -2.9 | 80.0 | 33.2 | **22.3** | 6.1 |
| | D1-PSNR | 73.7 | **86.4** | **60.1** | 63.5 | 80.8 | 65.6 | 71.7 | **71.7** | **87.3** | 39.6 | 23.1 | 50.7 |
| Point-to-Plane | D2 | 81.9 | 64.8 | 27.2 | 58.4 | 81.0 | 90.1 | -40.1 | 73.3 | 77.3 | -5.2 | 7.7 | 41.1 |
| | Hausdorff | 84.2 | 51.9 | 25.7 | 33.1 | 79.3 | 62.5 | 69.4 | 0.6 | 78.1 | 20.5 | 20.1 | -0.6 |
| | D2-PSNR | 79.2 | 84.0 | 46.7 | **64.2** | **86.5** | 51.9 | 61.4 | 69.4 | 81.8 | **53.9** | 20.5 | **55.1** |
| Plane-to-Plane | MSE | 72.0 | 42.8 | 57.9 | 45.4 | 53.0 | 65.5 | **77.4** | 36.0 | 33.1 | 22.4 | 23.1 | 29.6 |
| | RMS | 72.0 | 42.8 | 48.7 | 45.4 | 53.0 | 65.5 | **77.4** | 35.9 | 33.1 | 21.7 | 24.0 | 28.5 |
| | MEAN | 69.6 | 45.3 | 43.4 | 45.0 | 49.3 | 65.8 | 76.4 | 36.3 | 29.4 | 19.8 | 31.1 | 29.9 |
| - | No. Points | 55.8 | -11.4 | 17.9 | 18.2 | 59.9 | 38.2 | 55.2 | 55.3 | 63.4 | -28.4 | 8.8 | 13.1 |

### 5.6.1 Impact of Coding on the Point Cloud Quality Metric Correlation Performance

In the following impact of different characteristics of coding approaches (most importantly the number of decoded points) on state-of-the-art metrics, specifically Po2Po, Po2Pl and Pl2Pl metrics are investigated.

- **PCL Coding:** For PCL coded data, although D2-PSNR is the best performing metric, the Po2Po metrics have the best performance with high PLCC and SROCC for all rendering approaches, as shown in Table 5.7 and Table 5.8. As PCL controls the rate by reducing the number of decoded points, large objective errors and perceived distortions are visible for all renderings. This was expected since, when the compression ratio increases (lower rates), more and more points are discarded (due to octree pruning) and the remaining points are represented farther away from the original surface. PCL artefacts are strong enough to be visible even after the *RMesh* surface reconstruction.
- **G-PCC & V-PCC Coding:** As shown in Table 5.7 and Table 5.8, the objective quality metrics correlation performance for G-PCC is slightly lower (4 to 5% PLCC and SROCC for the best case) compared to PCL and shows the highest performance for Po2Po metrics (only for the *RPoint* and *RColor* sessions). The best performing metric for G-PCC is D1-PSNR for *RPoint* and D2-PSNR for *RColor* and *RMesh*. Moreover, none of the selected objective quality metrics performs well for V-PCC coded data. The selected metrics underestimate the similarity between the reference and decoded point clouds, especially for *RPoint* and *RMesh* renderings where the geometry errors are less visible, e.g., compared to *RColor*. Since both the G-PCC and V-PCC codecs tend to add points with respect to the reference point cloud (see Figure 5.9),

the density of points is increased and, thus, the perceived quality is higher (higher MOS). However, the objective quality metrics are not able to account for this effect and, thus, underperform for G-PCC and V-PCC codecs. In addition, since a wide range of values is obtained for the ratio of decoded over original number of points, notably depending on the codec (and also coding parameters), it is rather difficult to map errors to a perceptually meaningful metric; this makes the task of designing reliable objective quality metrics harder, especially when different types of codecs, with different coding artefacts, are jointly assessed ('All' column in Table 5.7 and Table 5.8). The objective quality metrics correlation performance for V-PCC is much lower compared to G-PCC (cf. Table 5.7 and Table 5.8). The projection-based V-PCC codec causes slight distortions on the geometry which are not very visible even for the lower bitrates. On the other hand, G-PCC artefacts are more visible, especially when the surface estimation (triangulation) process performs poorly.

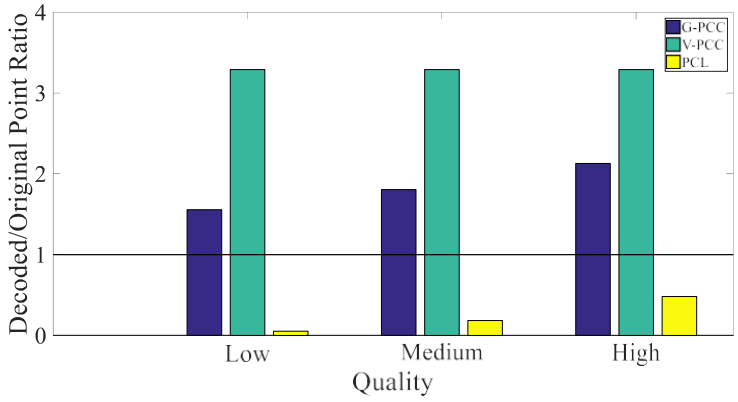



**Figure 5.9:** Average ratio of decoded over original number of points (1 means the reference and decoded number of points are the same).

Figure 5.10 shows the histogram of the square root of Po2Po for all the points in *Facade* and *Egyptian Mask* point clouds coded with G-PCC and V-PCC for which the same overall D1-PSNR (60 dB) was obtained. As shown, although the final D1-PSNR is the same, the error distribution is very different between G-PCC and V-PCC. V-PCC errors are closer to zero which makes them less perceptually visible while G-PCC errors have larger magnitudes and, thus, are more visible. This implies that G-PCC has a lower subjective similarity (MOS of 1.1) than V-PCC (MOS of 3.15) even when the D1-PSNR objective metric computes the same score (in this case 60 dB). This observation happens also for other objective quality metrics, such as D2-PSNR.

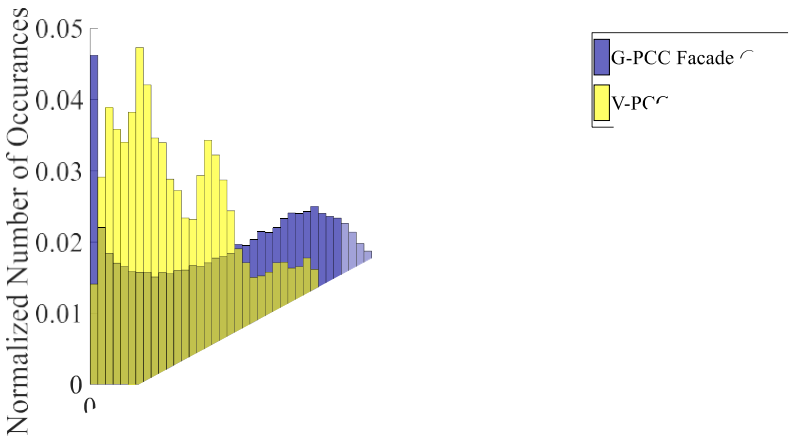



**Figure 5.10:** Po2Po error histograms for G-PCC and V-PCC codecs.

In summary, the objective quality metrics correlation performance highly depends on the coding distortions introduced, being satisfactory when point clouds are coded with PCL and G-PCC and performing poorly for V-PCC. Naturally, no objective quality metric performs well for all codecs together, a real problem when comparing the RD performance of very different coding paradigms.

### 5.6.2 Impact of Rendering on the Point Cloud Metrics Correlation Assessment

As previously concluded, rendering can significantly influence the visibility of coding artefacts (i.e., the perceived similarity/impairment between reference and decoded point clouds) and, thus, it is important to analyze the objective quality metrics correlation performance for different rendering solutions. The PLCC and SROCC value over *all data* for all sessions is rather low (less than 78.4% and 71.8% respectively) because the objective quality metrics cannot measure with accuracy all different types of distortions. However, as shown in Table 5.7 and Table 5.8, the best PLCC and SROCC values both occur for *RColor* rendering, for which higher MOS values were obtained. Thus, geometry quality metrics measure the perceived similarity better when color attributes are used.

The main reason is because subjects were able to better perceive degradations for medium and high quality ranges (which occur often with *RColor*, *cf.* Section 5.4.2) compared to low and medium quality ranges (which occur more often with *RPoint and RMesh*, *cf.* Section 5.4.2). For *RMesh* rendering, PLCC values are rather low comparing to the other rendering approaches, especially for the G-PCC and V-PCC codecs. For *RMesh*, point cloud data is converted to a polygonal mesh (surface reconstruction) for rendering and most the objective quality metrics have low correlation performance for this type of representation.

In summary, objective quality metrics account better for distortion artefacts and are more reliable when point-based rendering (with and without color, *RPoint* and *RColor*) is used to process the decoded point clouds before visualization.

### 5.6.3 Point Cloud Objective Metrics Correlation Assessment

In this section, correlation performance of the different metrics, sorted by their type, is evaluated for all stimuli in the subjective tests of this chapter and also for decoded point clouds of each codec individually:

**Po2Po metrics:** Po2Po metrics have a high PLCC performance for many cases but are especially better than others for the PCL and G-PCC codecs (*RColor* and *RPoint*). This is because PCL and G-PCC (to some extent) are octree-based point cloud codecs and, thus, some distortions still come from the positioning error related to the 3D partitioning of space into voxels, the target of this type of quality metrics. The Po2Po and Po2Pl Hausdorff metrics can also reach high PLCC values, especially for PCL data and the *RPoint* session (90.07). However, Hausdorff is not a very reliable quality metric when different types of coding distortions (all data and G-PCC/V-PCC) are considered together. The main reason is that only the maximum error is accounted and, thus, this type of metric is too sensible to outliers; this problem has been already observed in our previous work reported in Chapter 3.

**Po2Pl metrics:** Regarding Po2Pl metrics, the performance is very similar to Po2Po metrics, although slightly better for some cases since it considers the underlying surface from which the 3D point locations were sampled. Moreover, the D2-PSNR metric as a Po2Pl metric excels, being rather reliable and consistent for many cases, outperforming the corresponding D2 metric. The main reason is that the peak used (computed from the geometry coordinate precision) to convert MSE to PSNR values acts as an important normalizer.

**Pl2Pl metrics:** Pl2Pl metrics have, in general, worst PLCC performance when compared to Po2Po and Po2Pl metrics. This is mainly because it is rather difficult to obtain reliable normal vectors for the decoded point cloud, especially when some types of coding artefacts are present (e.g., holes) or the decoded point cloud is rather sparse [109]. However, these quality metrics seem to be the best choice for the V-PCC codec (for *RColor* and *RMesh* renderings) where geometry errors mainly come from coding artefacts in the 2D geometry maps and, thus, are more consistent among different parts of the point cloud.

As a curiosity, the number of decoded points could also be used as an objective quality metric, see last line of Table 5.7 and Table 5.8. As expected, this metric performs very poorly, especially for V-PCC data where the number of decoded points is typically larger than the number of original points and critically depends on some coding parameters, e.g., *B0* for the V-PCC occupancy maps.

### 5.6.4 Statistical Significance Analysis of Objective Metrics Performance

Besides the usual PLCC correlation evaluation, the difference in the performance of one objective quality metric with respect to another was assessed for statistically significance, using the procedure suggested in [110]. For that purpose, the so-called prediction residuals were first calculated by subtracting the subjective scores from the predicted subjective values, obtained by applying a nonlinear logistic fitting to the objective quality scores. These prediction residuals were

obtained for every point cloud objective quality metric. Then, the one-tailed F-test was applied to the prediction residuals, to assess if the difference in PLCC performance between any two given point cloud objective quality metrics is statistically significant at some significance level. In general, the significance level should be set based on the sample size (cardinality) being evaluated for an increased power of the test (i.e., probability of rejecting the null hypothesis when it is not true) [111]; since the cardinality of the prediction residuals is 18 for a single point cloud codec (PCL, G-PCC and V-PCC), a 0.2 significance level was used [111]. The F-test assumes that the samples are normally distributed, and thus the kurtosis test was used to verify whether all prediction residuals followed a Gaussian distribution, which was the case for all the objective metrics, except the No. Points, which is recognized as a very poor quality metric.

In this work, the F-test null hypothesis is that the prediction residuals for the two objective quality metrics being compared are obtained from normal distributions with the same variance, which means that the pair of objective metrics under evaluation is statistically similar. The alternative hypothesis is that the prediction residuals for the two objective quality metrics being compared are obtained from normal distributions with different variances, which means that the pair of objective metrics under evaluation are statistically different. By computing the ratio between the variances of the two prediction residuals, the test statistic, $F$, was obtained, which was then compared to the F-test critical value, $F_{critical}$; the F-test critical value depends on the significance level and the sample sizes. When $F$ is larger than $F_{critical}$, then the null hypothesis can be rejected, which means that the objective quality metrics under evaluation are statistically different; otherwise, the null hypothesis cannot be rejected, meaning that the objective metrics under evaluation are statistically indistinguishable. Since the test statistic $F$ is always computed with the objective quality metric with larger prediction residual variance in the numerator, the objective quality metric in the denominator corresponds to the metric with the best performance whenever the null hypothesis is rejected.

The F-test results are presented in Table 5.9,

Table 5.10 and Table 5.11 for each of the test sessions. for *RPoint*, *RColor* and *RMesh* rendering, respectively. Each table entry is the codeword that represents the F-test outcome for PCL, G-PCC, V-PCC and All codecs, respectively; a ‘‘1’’, ‘‘_’’, or ‘‘0’’ means that the performance of the objective metric in a given row (in terms of the correlation of its predictions with MOS) is better, similar, or worse than the performance of the objective metric in a given column, respectively. The results obtained confirm that Po2Po and Po2Pl metrics have the best overall performance for many coding scenarios (PCL, G-PCC and for ‘All’ codecs), especially the PSNR based metrics which are consistently better than D1 and Hausdorff based metrics. Between the D1 and D2 PSNR, there is no statistical difference, which means that both metrics achieve similar correlation performance. Moreover, Pl2Pl metrics have the best performance for V-PCC decoded data (for *RColor*

rendering). In summary, the statistical significance results allow confirming that the conclusions drawn in Sections 5.6.1 to 5.6.3 before, regarding the superiority of a quality metric are valid.

**Table 5.9:** Statistical significance test results for the selected point cloud geometry objective quality metrics with *RPoint* rendering. Each entry is the codeword representing the best outcome for the PCL, G-PCC, V-PCC and 'All' codecs.

| | | **Po2Po** | | | **Po2Pl** | | | **Pl2Pl** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **D1** | **Hausdorff** | **D1-PSNR** | **D2** | **Hausdorff** | **D2-PSNR** | **MAD** | **MSAD** | **RMSAD** |
| **Po2Po** | **D1** | ---- | 1--1 | 00-0 | ---- | 1--- | 00-0 | ---- | ---- | ---- |
| | **Hausdorff** | 0--0 | ---- | 00-0 | 0--- | ---- | 00-0 | 0--0 | 0--0 | 0--0 |
| | **D1-PSNR** | 11-1 | 11-1 | ---- | 11-1 | 11-1 | ---- | 11-1 | 11-1 | 11-1 |
| **Po2Pl** | **D2** | ---- | 1--- | 00-0 | ---- | 1--- | 00-0 | ---- | ---- | ---- |
| | **Hausdorff** | 0--- | ---- | 00-0 | 0--- | ---- | 00-0 | 0--- | 0--- | 0--- |
| | **D2-PSNR** | 11-1 | 11-1 | ---- | 11-1 | 11-1 | ---- | 11-1 | 11-1 | 11-1 |
| **Pl2Pl** | **MAD** | ---- | 1--1 | 00-0 | ---- | 1--- | 00-0 | ---- | ---- | ---- |
| | **MSAD** | ---- | 1--1 | 00-0 | ---- | 1--- | 00-0 | ---- | ---- | ---- |
| | **RMSAD** | ---- | 1--1 | 00-0 | ---- | 1--- | 00-0 | ---- | ---- | ---- |

**Table 5.10:** Statistical significance test results for the selected point cloud objective quality metrics with *RColor* rendering. Each entry is the codeword representing the best outcome for the PCL, G-PCC, V-PCC and 'All' codecs.

| | | **Po2Po** | | | **Po2Pl** | | | **Pl2Pl** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **D1** | **Hausdorff** | **D1-PSNR** | **D2** | **Hausdorff** | **D2-PSNR** | **MAD** | **MSAD** | **RMSAD** |
| **Po2Po** | **D1** | ---- | 11-1 | 01-0 | ---- | 1101 | 01-0 | 1101 | 1101 | 1101 |
| | **Hausdorff** | 00-0 | ---- | 0-10 | 00-0 | ---- | 0--0 | --0- | --0- | --0- |
| | **D1-PSNR** | 10-1 | 1-01 | ---- | 1--1 | 1-01 | ---- | 1-01 | 1-01 | 1-01 |
| **Po2Pl** | **D2** | ---- | 11-1 | 0--0 | ---- | 1-01 | 01-0 | 1-01 | 1-01 | 1-01 |
| | **Hausdorff** | 0010 | ---- | 0-10 | 0-10 | ---- | 0--0 | ---- | ---- | ---- |
| | **D2-PSNR** | 10-1 | 1--1 | ---- | 10-1 | 1--1 | ---- | 1-01 | 1-01 | 1-01 |
| **Pl2Pl** | **MAD** | 0010 | --1- | 0-10 | 0-10 | ---- | 0-10 | ---- | ---- | ---- |
| | **MSAD** | 0010 | --1- | 0-10 | 0-10 | ---- | 0-10 | ---- | ---- | ---- |
| | **RMSAD** | 0010 | --1- | 0-10 | 0-10 | ---- | 0-10 | ---- | ---- | ---- |

**Table 5.11:** Statistical significance test results for the selected point cloud objective quality metrics with *RMesh* rendering. Each entry is the codeword representing the best outcome for the PCL, G-PCC, V-PCC and 'All' codecs.

| | | **Po2Po** | | | **Po2Pl** | | | **Pl2Pl** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **D1** | **Hausdorff** | **D1-PSNR** | **D2** | **Hausdorff** | **D2-PSNR** | **MAD** | **MSAD** | **RMSAD** |
| **Po2Po** | **D1** | ---- | 10-- | 00-0 | ---- | 10-- | 00-0 | 1--- | 1--- | 1--- |
| | **Hausdorff** | 01-- | ---- | 0--0 | 0--- | ---- | 0--0 | ---- | ---- | ---- |
| | **D1-PSNR** | 11-1 | 1--1 | ---- | 11-1 | 1--1 | ---- | 1--1 | 1--1 | 1--1 |
| **Po2Pl** | **D2** | ---- | 1--- | 00-0 | ---- | ---- | 00-0 | ---- | ---- | ---- |
| | **Hausdorff** | 01-- | ---- | 0--0 | ---- | ---- | 0--0 | ---- | ---- | ---- |
| | **D2-PSNR** | 11-1 | 1--1 | ---- | 11-1 | 1--1 | ---- | 1--1 | 1--1 | 1--1 |
| **Pl2Pl** | **MAD** | 0--- | ---- | 0--0 | ---- | ---- | 0--0 | ---- | ---- | ---- |
| | **MSAD** | 0--- | ---- | 0--0 | ---- | ---- | 0--0 | ---- | ---- | ---- |
| | **RMSAD** | 0--- | ---- | 0--0 | ---- | ---- | 0--0 | ---- | ---- | ---- |

In summary, when all codecs and all renderings are considered, the D1 and D2 PSNR objective quality metrics have the highest correlation with subjective scores. These metrics correspond to the ones previously selected by MPEG for the PCC Call for Proposals and currently used in the MPEG Common Test Conditions [104]. To the best of the authors knowledge, this is the first time that the selection of these quality metrics has been validated with MOS obtained using a well-defined procedure. From the results presented in this work, there is still significant room for improvement, especially if the goal is to achieve the same level of correlation performance that image objective quality metrics (e.g., SSIM based) have achieved for 2D image and video coding.

## 5.7 Final Remarks

The main objectives of this chapter were to study the impact of the rendering process on the perceived quality of decoded point clouds and the correlation performance of available point cloud geometry objective metrics. To achieve these objectives, three representative point cloud coding solutions and three point cloud rendering solutions were selected as well as a wide set of objective quality metrics. The subjective experiments suggest that geometry coding distortions can be masked by using the color attributes and (to a less extent) by surface reconstruction methods. Moreover, point cloud codecs produce distinct coding artefacts that have different impacts in terms of the final perceived quality, e.g., for PCL decoded data, geometry distortions are clearly visible for all rendering methods. Regarding the objective quality metrics correlation evaluation, the results show that a careful selection of the objective metrics is necessary to have a reliable measure of the decoded point clouds quality. Also, for some codecs and rendering solutions, the current metrics are not very reliable, e.g., for V-PCC coded data; this is rather critical since V-PCC is expected to become the first coding standard to be deployed in the market. Moreover, some of the

objective quality metrics have a rather limited scope with significantly degraded accuracy, for some specific rendering solutions. For example, Hausdorff performance is only acceptable for PCL codec.

The work in this chapter has led to a journal publication, notably:
**A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Point Cloud Rendering after Coding: Impacts on Subjective and Objective Quality," *IEEE Transactions on Multimedia*, November 2020.

# Part III. Objective Quality Assessment

# Chapter 6

# Generalized Hausdorff Distance-based PSNR Quality Metric for Point Cloud Geometry

## 6.1 Context and Objectives

As mentioned in Chapter 4, the PSNR computed from the classical Hausdorff distance shows a high objective-subjective correlation performance for octree-based point cloud codecs [112] [25]. However, MPEG no longer uses this metric for point cloud geometry quality assessment due to its low reliability for the emerging MPEG point cloud coding standards, notably Geometry-based Point Cloud Compression (G-PCC) [113] and Video-based Point Cloud Compression (V-PCC) [103], which are based on other coding paradigms. In fact, the classical Hausdorff distance-based PSNR quality metric for point cloud geometry is very sensitive to outlier points in the decoded point cloud, which may not even be visible (or rendered), as the Hausdorff distance corresponds to the largest of all distances from a point in one point cloud to the closest point in the other point cloud (reference to decoded, and vice-versa). This extreme outlier sensitive behavior was the motivation to perform a statistical analysis of the distances' distribution for point clouds coded with different codecs, to find a more reliable quality metric for point cloud geometry. In this context, the objective of this work was to design and evaluate a more reliable point cloud geometry quality metric, notably in terms of correlation performance, which is based on the so-called Generalized Hausdorff (GH) distance [114], an extension of the classical Hausdorff distance.

The generalized Hausdorff distance measures the distance using the $K^{th}$ ranked distance instead of the last ranked (i.e., maximum) distance; this is equivalent to the last ranked (maximum) distance in a ($K$-values) limited portion of the sorted distances. While it has been previously used for object matching in noisy images [115], the generalized Hausdorff distance has never been used for point cloud quality assessment. Since a small number of points with large errors/distances (i.e., outlier

points) dispersed in a decoded point cloud may not be even very visible, the generalized Hausdorff distance may be a better choice for point cloud geometry quality assessment since outlier points may be discarded, and their limited quality impact avoided.

The rest of this chapter is organized as follows. Section 6.2 describes the proposed point cloud geometry quality metric. The experimental results are presented and analyzed in Section 06.3. Finally, a summary of the conclusions and final remarks are presented in Section 6.4.

## 6.2 Proposed Point Cloud Geometry Quality Metric

The classical Hausdorff distance is very sensitive to outliers, since one or more points with a large error magnitude will dominate the final quality score even for cases where these points are not much visible. The classical Hausdorff distance defined in Section 2.6.3 can be generalized to compute the distance over a subset of data/distances after ranking all the distance/error values. Instead of taking the maximum distance over all the distances as in the classical Hausdorff distance, the generalized Hausdorff distance for rank $K$ is computed using only the $K$ lowest distance values after ranking all the distances in ascending order. Thus, the $K^{th}$ ranked generalized Hausdorff distance is defined as:

$$d_{GH-K}(A,B) = {}^{per}K^{th}_{a\in A}\ d(a,B) \tag{6.1}$$

where ${}^{per}K^{th}_{a\in A}$ is the $K^{th}$ ranked distance such that $(K/N_A)\times 100 = per\%$ and $N_A$ is the total number of points in point cloud $A$. For example, the 480th ranked distance in a point cloud with 600 points is the maximum distance obtained from the $per$ = (480 / 600) × 100 = 80% lowest distance values, after sorting all the distances in ascending order. Note that in this context, the $d$ distance corresponds to the squared Euclidean distance from a point in one point cloud to the nearest neighbor on the other point cloud (Po2Po distance), which can be naturally extended to account for the normal vector at each point (Po2Pl distance). These distances are described in Section 2.6.3.

In the experiments performed to design a better quality metric, 15 directed generalized Hausdorff distances (see below which ones) were obtained by assigning different values to *per%* in the generalized Hausdorff distance as defined in (6.1), where $d_{100/Na}$ and $d_{100}$ are special cases corresponding to:

$$d_{\underline{100}}(A,B) = \min_{a\in A} d(a,B) = {}^{N_A}K^{th}_{a\in A}\ d(a,B) \tag{6.2}$$
$$N_a$$

$$d_{100}(A,B) = \max_{a\in A} d(a,B) = {}^{100}K^{th}_{a\in A}\ d(a,B) \tag{6.3}$$

Note that $d_{100}$ corresponds to the classical Hausdorff distance mentioned before. Moreover, four different pooling functions are considered to compute the undirected distance (i.e., considering both directions, i.e., reference to decoded and vice-versa) between two point clouds, notably:

$$pool_{min}(d(A,B),d(B,A)) = \min(d(A,B),d(B,A)) \tag{6.4}$$

$$pool_{max}(d(A,B),d(B,A)) = \max(d(A,B),d(B,A)) \tag{6.5}$$

$$pool_{avg}(d(A,B),d(B,A)) = \frac{d(A,B)+d(B,A)}{2} \tag{6.6}$$

$$pool_{wavg}(d(A,B),d(B,A)) = \frac{N_A d(A,B)+N_B d(B,A)}{N_A+N_B} \tag{6.7}$$

By combining the 15 directed distances with these four proposed undirected distance pooling functions, 60 undirected distances were obtained, labelled in the following as $D_{per,pool}$, which will be used to measure the distortion between two point clouds (reference and decoded). After, the corresponding PSNR quality metric was computed for all these undirected distances, using them as a substitute to $d_{MSE}$ equation (6.8) while using the same $p$.

$$PSNR = 10\log_{10}\frac{3p^2}{MSE} \tag{6.8}$$

$$PSNR_{perpool} = 10\log_{10}\frac{3p^2}{D_{perpool}} \tag{6.9}$$

where $p$ is the precision of point cloud. The full list of these 60 PSNR quality metrics computed from the 60 undirected distances, $D_1$ to $D_{60}$, is shown in Table 6.1 where $N_A$ is the number of points in point cloud $A$. For example, $PSNR_{65,wavg}$ is obtained from the undirected distance $D_{65,wavg}$, which results from applying the weighted average pooling function to pool the directed distances $d_{65}$ for both directions associated to $per = 65\%$; note that $PSNR_{100,max}$ corresponds to the classical Hausdorff distance-based PSNR. In the following, all these PSNR quality metrics will be assessed to identify the best performing ones.

## 6.3 Performance Evaluation

This section assesses the correlation performance of the proposed point cloud geometry quality metrics for different ranks and pooling functions. This assessment allows to find the best performing quality metric for point clouds with artifacts generated by the MPEG standard point cloud codecs described in Section 5.3.

### 6.3.1 Subjective Evaluation Dataset

In this Chapter, the MOS scores and corresponding point clouds available in the *IST Rendered Point Cloud (IRPC)* quality assessment dataset presented in Chapter 5 are used, since they correspond to very extensive and representative conditions.

The reference and decoded point clouds were compared and evaluated by subjects using a DSIS subjective assessment protocol. The adopted point cloud dataset includes six point clouds from the

MPEG repository [104], notably *Egyptian Mask* and *Frog* from the inanimate objects category; *Facade9* and *House without roof* from facades and buildings; and *Loot* and *Longdress* from people. The key characteristics of these point clouds are listed in Table 6.2.

**Table 6.1:** Generalized Hausdorff distance-based PSNR metrics obtained from 60 undirected distances, derived from 15 directed distances (in rows) and 4 pooling functions (in columns).

| | ***per*% in (8)** | **$pool_{min}$** | **$pool_{max}$** | **$pool_{avg}$** | **$pool_{wavg}$** |
|---|---|---|---|---|---|
| **$d_1$** | ***100/$N_A$*** | $PSNR_{100/NA,min}$ | $PSNR_{100/NA,max}$ | $PSNR_{100/NA,avg}$ | $PSNR_{100/NA,wavg}$ |
| **$d_{50}$** | ***50*** | $PSNR_{50,min}$ | $PSNR_{50,max}$ | $PSNR_{50,\ avg}$ | $PSNR_{50,\ wavg}$ |
| **$d_{60}$** | ***60*** | $PSNR_{60,min}$ | $PSNR_{60,max}$ | $PSNR_{60,\ avg}$ | $PSNR_{60,\ wavg}$ |
| **$d_{65}$** | ***65*** | $PSNR_{65,min}$ | $PSNR_{65,max}$ | $PSNR_{65,\ avg}$ | $PSNR_{65,\ wavg}$ |
| **$d_{70}$** | ***70*** | $PSNR_{70,min}$ | $PSNR_{70,max}$ | $PSNR_{70,\ avg}$ | $PSNR_{70,\ wavg}$ |
| **$d_{75}$** | ***75*** | $PSNR_{75,min}$ | $PSNR_{75,max}$ | $PSNR_{75,\ avg}$ | $PSNR_{75,\ wavg}$ |
| **$d_{80}$** | ***80*** | $PSNR_{80,min}$ | $PSNR_{80,max}$ | $PSNR_{80,\ avg}$ | $PSNR_{80,\ wavg}$ |
| **$d_{85}$** | ***85*** | $PSNR_{85,min}$ | $PSNR_{85,max}$ | $PSNR_{85,\ avg}$ | $PSNR_{85,\ wavg}$ |
| **$d_{90}$** | ***90*** | $PSNR_{90,min}$ | $PSNR_{90,max}$ | $PSNR_{90,\ avg}$ | $PSNR_{90,\ wavg}$ |
| **$d_{95}$** | ***95*** | $PSNR_{95,min}$ | $PSNR_{95,max}$ | $PSNR_{95,\ avg}$ | $PSNR_{95,\ wavg}$ |
| **$d_{96}$** | ***96*** | $PSNR_{96,min}$ | $PSNR_{96,max}$ | $PSNR_{96,\ avg}$ | $PSNR_{96,\ wavg}$ |
| **$d_{97}$** | ***97*** | $PSNR_{97,min}$ | $PSNR_{97,max}$ | $PSNR_{97,\ avg}$ | $PSNR_{97,\ wavg}$ |
| **$d_{98}$** | ***98*** | $PSNR_{98,min}$ | $PSNR_{98,max}$ | $PSNR_{98,\ avg}$ | $PSNR_{98,\ wavg}$ |
| **$d_{99}$** | ***99*** | $PSNR_{99,min}$ | $PSNR_{99,max}$ | $PSNR_{99,\ avg}$ | $PSNR_{99,\ wavg}$ |
| **$d_{100}$** | ***100*** | $PSNR_{100,min}$ | $PSNR_{100,max}$ | $PSNR_{100,\ avg}$ | $PSNR_{100,\ wavg}$ |

**Table 6.2:** Test materials and respective characteristics.

| **Point Cloud Name** | **No. Points** | **Precision** | **Category** | **Signal Peak (*p*)** |
|---|---|---|---|---|
| *Egyptian Mask* | 272,684 | 12-bit | Inanimate Objects | 4095 |
| *Frog* | 3,614,251 | 12-bit | Inanimate Objects | 4095 |
| *Facade9* | 1,596,085 | 12-bit | Facades & Buildings | 4095 |
| *House without roof* | 4,848,745 | 12-bit | Facades & Buildings | 4095 |
| *Loot* | 805,285 | 10-bit | People | 1023 |
| *Longdress* | 857,966 | 10-bit | People | 1023 |

The selected point clouds have been coded with three rates/qualities using the following point cloud codecs: i) octree-based codec from PCL [89]; ii) MPEG G-PCC; and iii) MPEG V-PCC codec. The selected codecs represent three different coding paradigms, notably PCL for tree structures, MPEG G-PCC for surface models (when *Trisoup* is used) and MPEG V-PCC for projection-based coding (see Section 2.3 for details on these codecs). The bitrates for each codec were selected from the rates defined in the MPEG Common Test Conditions (CTC) [104], to have three distinguishable qualities for each point cloud. The coding parameters used for the PCL, MPEG G-PCC and V-PCC codecs are the same as in Chapter 5.

The decoded point clouds have been subjectively assessed using a DSIS subjective assessment protocol in three sessions, each corresponding to a different rendering approach, notably *RPoint* (rendering using a point representation with uniform color and shading), *RMesh* (rendering using a mesh representation with uniform color and shading), and *RColor* (rendering using a point representation and original colors). Because there were no coloring and interpolation processes involved in *RPoint* rendering, geometry coding artifacts are less masked in this rendering approach (more details about this dataset are presented in Chapter 5). Since the proposed point cloud quality metrics defined in Section 6.2 only evaluate point cloud geometry, only the *RPoint* session MOS [105] will be used in the following. There were 20 subjects participating in each assessment session and no outliers were found for the *RPoint* session, whose MOS is used in the experiments presented next.

### 6.3.2 Experimental Results

As for previous chapters, the correlation performance of a quality metric is measured by computing the correlation of the metric scores with respect to human expressed MOS. For point cloud geometry quality assessment, this implies that, even when all points are not exactly in the same positions in the two point clouds *A* and *B* under comparison, as long as *A* and *B* look perceptually similar, a good metric should create an error close to zero or the maximum score. Thus, the proposed quality metric should discard distances/errors that are not visible, i.e., perceptually relevant, to approximate this thresholding effect of the human visual system. Moreover, to obtain reliable objective quality assessment metrics, the adopted distances need to discriminate well between different perceptual quality levels for point cloud geometry.

Several experiments with decoded point clouds using the three previously selected point cloud codecs have shown $D_{100/N,-}$ and $D_{50,-}$ ( where - means any pooling) are 0 most of the time, especially for high qualities. Moreover, $D_{60,min}$, $D_{65,min}$ and $D_{70,min}$ are also often 0 for V-PCC coding, clearly indicating that they do not have enough discriminatory power to evaluate V-PCC decoded point clouds. Thus, $D_{100/N,-}$ to $D_{50,-}$ are not further used in the following performance assessment exercise, which targets identifying the best performing (PSNR) quality metric and the corresponding generalized Hausdorff distance rank.

To evaluate the objective-subjective correlation performance for the proposed point cloud geometry quality metrics and associated distances, non-linear cubic regression has been used to

map the computed objective quality scores (PSNR) into the MOS values available from the DSIS subjective test [105]. In this case, the following cubic function was used:

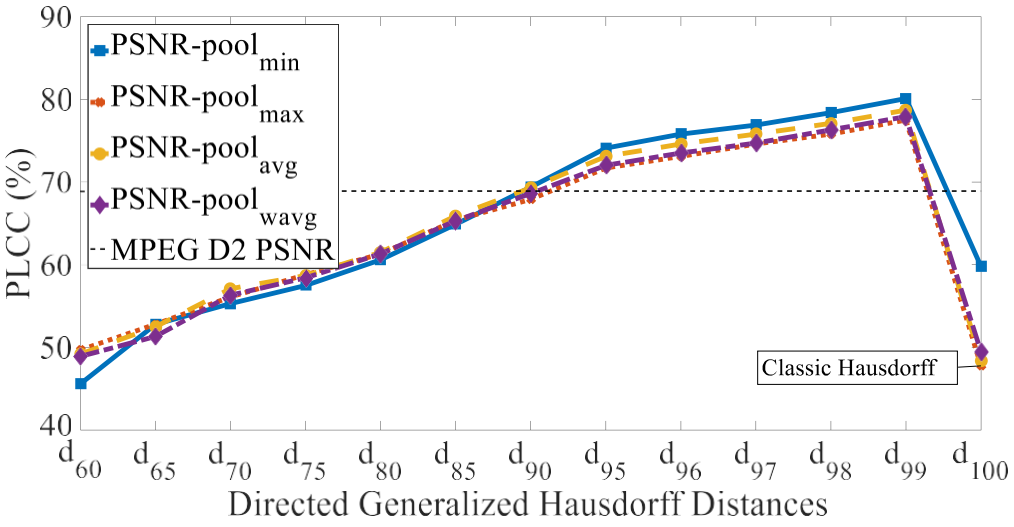
$$MOS_p = a + by + cy^2 + dy^3 \quad (6.8)$$

where $y$ are the objective metric scores and $a$, …, $d$ are the regression model parameters [33]. PLCC, SROCC, and RMSE, were again selected as the correlation performance metrics as for previous chapters, to assess the objective-subjective correlation performance.

Figure 6.1 shows the overall PLCC performance for the Po2Po and Po2Pl PSNR computed for the generalized Hausdorff distance using different ranks and pooling functions in comparison with the MPEG D1-PSNR and MPEG D2-PSNR metrics by considering the MOS scores from all selected codecs together. The results show that above *per = 90%* ($d_{90}$), the generalized Hausdorff distance-based PSNR metric PLCC starts to outperform the MPEG D1-PSNR and D2-PSNR PLCC when the three selected point cloud codecs are considered together. There is also a significant drop in PLCC from the generalized Hausdorff-based PSNR to the classical Hausdorff-based PSNR (extreme right side of the charts), thus highlighting that a small number of dispersed large distances/errors (outliers) are not visible since they did not impact the MOS values in the same way; in fact, the PLCC charts show that their consideration on the objective quality metrics largely penalizes the objective-subjective correlation. The best PLCC performing Po2Po metric is $PSNR_{98,avg}$, corresponding to $D_{98,avg}$, which considers 98% of the data/distances; for Po2Pl metrics, the best performing metric is $PSNR_{99,min}$, corresponding to $D_{99,min,}$ where 99% of data/distances are used. Although the experimental results do not show a significant PLCC difference between the various pooling functions, for Po2Po distances, the $pool_{min}$ function is clearly the best for lower *per%* while the $pool_{avg}$ function is the best for higher *per%*. For Po2Pl distances, the opposite behavior can be observed.

The PLCC results in Figure 6.1 for *per* = 100% ($d_{100}$) clearly indicate the presence of outliers in the distances' distribution, which strongly penalizes the correlation. Figure 6.2 shows the ranked distance for different percentage of data/distances (*per* in equation (6.1)), individually for the PCL, G-PCC and V-PCC codecs. The top row in Figure 6.2 shows the distances from reference to decoded point clouds while the bottom row shows the distances from decoded to reference point clouds. For G-PCC and V-PCC, there is a large and sudden increase of the maximum distance, close to *per* = 100% while, for PCL, the distances increase much slower. This behavior highlights that, for G-PCC and V-PCC, the decoded point clouds have a very small portion of distances/errors much larger than the average distance/error. This behavior also explains why the classical Hausdorff distance-based metrics performs well for PCL [112] [25] but not for G-PCC and V-PCC, thus justifying MPEG not adopting this quality metric anymore.

(a)




(b)

**Figure 6.1:** PLCC performance for (a) Po2Po and (b) Po2Pl generalized Hausdorff distance-based PSNR for different poolings as a function of the directed rank distance in comparison with the MPEG D1 PSNR and D2 PSNR quality metrics.

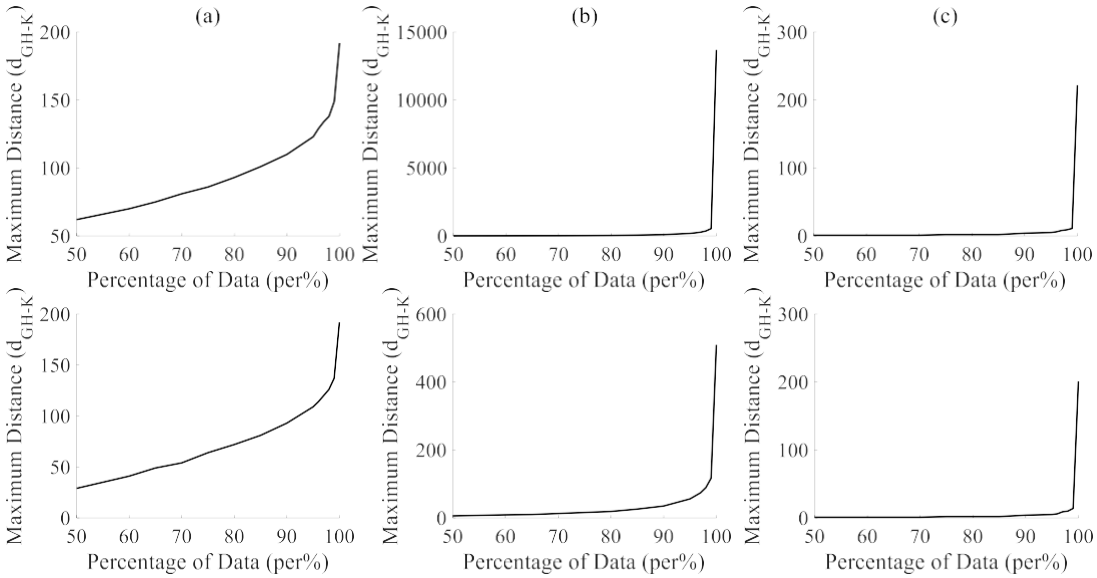



**Figure 6.2:** Maximum of directed squared distances from reference to decoded (top) and decoded to reference (bottom) for each selected percentage of data for the three geometry codecs: (a) PCL; (b) G-PCC; (c) V-PCC.

To assess the proposed quality metrics correlation improvements regarding the MPEG metrics for each of the selected point cloud codecs, Figure 6.3 shows the PLCC between the MOS values and the Po2Po generalized Hausdorff-based PSNR metric with max pooling, before applying the fitting function in equation (6.8).

For PCL, the generalized Hausdorff distance-based PSNR metric correlation performance with different percentages of data is similar to MPEG D1-PSNR, which considers all distances/errors. This means that the errors introduced by PCL coding do not include a significant number of outliers. However, for the other two codecs, by not considering some of the distances (i.e., outliers), large correlation performance improvements can be achieved, notably for $d_{95}$ to $d_{98}$.

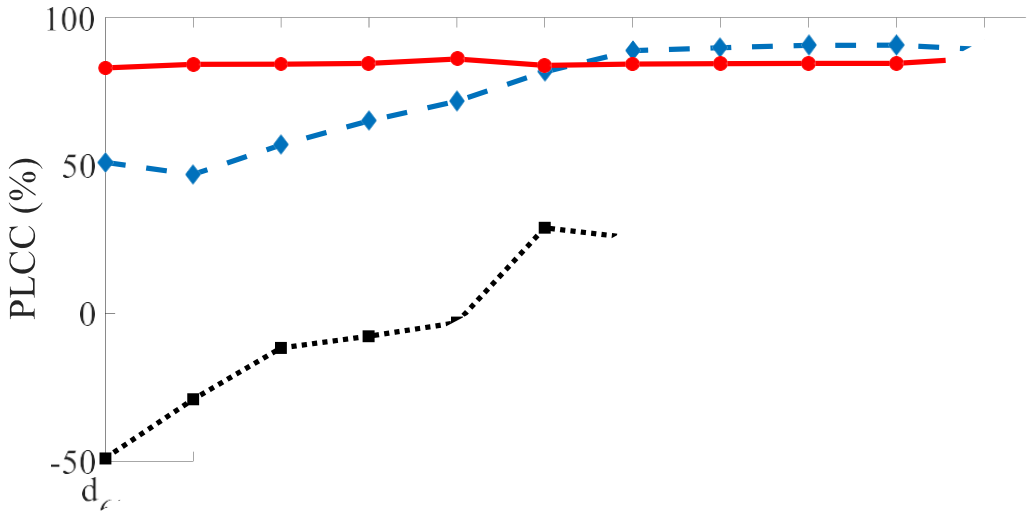



**Figure 6.3:** PLCC performance for Po2Po generalized Hausdorff PSNR using maximum function pooling and D1 PSNR for the three selected point cloud codecs.

Figure 6.4 shows the distribution of the coding errors for the *Egyptian Mask* point cloud coded with PCL, G-PCC and V-PCC. As it can be seen from Figure 6.4 considering that warmer colors show larger magnitude of error in the range of error vectors associated to decoded point clouds of each codec, the G-PCC and V-PCC decoded point clouds have a very small number of errors whose magnitude is much larger than the average error magnitude (being even invisible at naked eye) and those errors are dispersed over the decoded point cloud. Regarding the PCL decoded point cloud, most of the error magnitudes are around the average error magnitude, which, once more, indicates that the number of outliers (i.e., large errors) introduced by PCL coding is small.

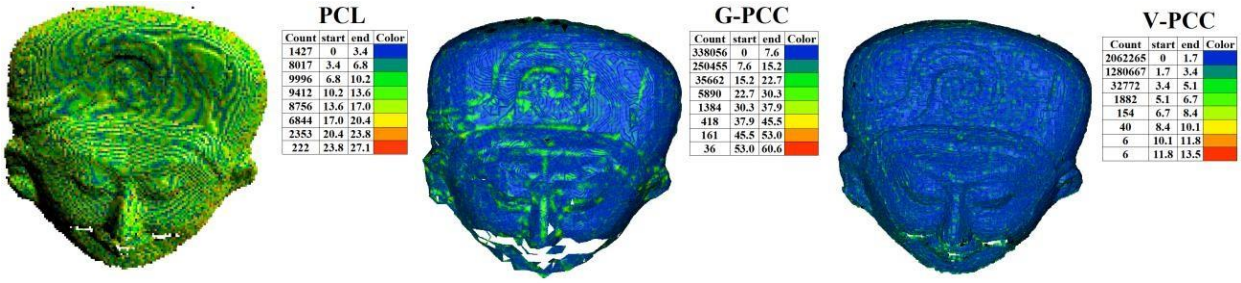



**Figure 6.4:** Distribution of coding errors throughout the decoded *Egyptian Mask* point cloud for PCL, G-PCC and V-PCC codecs. The tables show the number of error occurrences within a certain range, to which a specific color is assigned.

**Table 6.3:** PLCC, SROCC and RMSE performance for the best generalized Hausdorff distance-based PSNR quality metric for each point cloud codec individually and All codecs together.

| | | **Po2Po metrics** | | | | **Po2Pl metrics** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | ***Quality metric*** | ***PLCC*** | ***SROCC*** | ***RMSE*** | ***Quality metric*** | ***PLCC*** | ***SROCC*** | ***RMSE*** |
| **PCL** | ***Best GH*** | $PSNR_{100,max}$ | **89.5** | **79.9** | **0.5** | $PSNR_{85,max}$ | **90.5** | **86.7** | **0.5** |
| | ***MPEG*** | D1-PSNR | 87.0 | 73.8 | 0.6 | D2-PSNR | 90.2 | 85.8 | 0.6 |
| | ***Gain*** | - | **2.5** | **6.1** | **0.1** | - | **0.3** | **0.9** | **0.1** |
| **G-PCC** | ***Best GH*** | $PSNR_{98,max}$ | **91.3** | **93.5** | **0.4** | $PSNR_{97,min}$ | **89.4** | 91.7 | **0.4** |
| | ***MPEG*** | D1-PSNR | 86.9 | 87.4 | 0.5 | D2-PSNR | 89.1 | **93.0** | 0.5 |
| | ***Gain*** | - | **4.4** | **6.1** | **0.1** | - | **0.3** | -1.3 | **0.1** |
| **V-PCC** | ***Best GH*** | $PSNR_{98,min}$ | **69.5** | **67.6** | **0.5** | $PSNR_{96,min}$ | **76.6** | **79.9** | **0.4** |
| | ***MPEG*** | D1-PSNR | 53.1 | 62.0 | 0.6 | D2-PSNR | 51.4 | 49.6 | 0.6 |
| | ***Gain*** | - | **16.4** | **5.6** | **0.1** | - | **25.2** | **30.3** | **0.2** |
| **All Codecs** | ***Best GH*** | $PSNR_{98,avg}$ | **79.1** | **77.9** | **0.6** | $PSNR_{99,min}$ | **80.1** | **77.7** | **0.6** |
| | ***MPEG*** | D1-PSNR | 67.3 | 64.7 | 0.7 | D2-PSNR | 68.9 | 65.1 | 0.7 |
| | ***Gain*** | - | **11.8** | **13.2** | **0.1** | - | **11.2** | **12.6** | **0.1** |

Finally, the PLCC and SROCC performances for the best performing generalized Hausdorff (GH) based PSNR quality metrics, considering each codec individually and altogether, are shown in Table 6.3. These results allow concluding:

- **PCL** – For PCL, the Po2Po generalized Hausdorff-based PSNR with rank 100% (classical Hausdorff) outperforms MPEG D1-PSNR by 2.5%, 6.1% and 0.1 for PLCC, SROCC and RMSE, respectively; a similar behavior happens for the Po2Pl generalized Hausdorff with rank 85% ($PSNR_{85,max}$) with gains of 0.3%, 0.9% and 0.1 for PLCC, SROCC and RMSE, respectively, over MPEG D2 PSNR.
- **G-PCC** – For G-PCC, the best performing quality metric is the Po2Po generalized Hausdorff-based with rank 98% PSNR metric ($PSNR_{98,max}$) with gains of 4.4%, 6.1% and 0.1 for PLCC, SROCC and RMSE, respectively, over MPEG D1-PSNR. For Po2Pl, the generalized Hausdorff-based PSNR slightly outperforms MPEG D2-PSNR except for SROCC, where a small loss occurs.
- **V-PCC** – For V-PCC, there are major correlation gains compared to the MPEG quality metrics. The best quality metric for V-PCC is the Po2Pl generalized Hausdorff-based with rank 96% based PSNR metric ($PSNR_{96,min}$) with gains of 25.2%, 30.3% and 0.2 for PLCC, SROCC and RMSE, respectively, over MPEG D2-PSNR. Large gains also occur for the Po2Po

generalized Hausdorff-based with rank 90% PSNR metric ($PSNR_{98,min}$) over MPEG D1-PSNR. This allows to conclude that the filtering of outliers is more important for this point cloud codec, especially since a large number of points is added during the point cloud decoding process.

- **All codecs** – As already shown in Figure 6.1, the generalized Hausdorff based PSNR with ranks from 90% to 99% outperform the MPEG quality metrics when decoded data from all codecs are considered together. More precisely, Table 6.3 shows that, considering all codecs, the best Po2Po metric is the generalized Hausdorff-based with rank 98% PSNR metric ($PSNR_{98,avg}$) with gains of 11.8%, 13.2% and 0.1 for PLCC, SROCC and RMSE, respectively, over MPEG D1-PSNR; the best Po2Pl metric is the generalized Hausdorff-based with rank 99% PSNR metric ($PSNR_{99,min}$) with gains of 11.2%, 12.6% and 0.1 for PLCC, SROCC and RMSE, respectively, over MPEG D2 PSNR.

As shown in Table 6.3, the generalized Hausdorff distance-based PSNR outperforms the MPEG D1 PSNR and D2 PSNR metrics, if the right rank (i.e. *per%*) is used. In general, for Po2Po distances, except for data coded with PCL, generalized Hausdorff with 98% of data ($d_{98}$) has the best correlation performance. For Po2Pl metrics, the best generalized Hausdorff-based metric varies for each codec. Although not shown in Table 6.3, the PSNR associated with classical Hausdorff distance ($PSNR_{100,max}$) shows poor correlation with MOS values, except for PCL.

**Table 6.4:** Correlation performance for the best overall generalized Hausdorff distance-based PSNR quality metric.

| | **Po2Po metrics** | | | | | **Po2Pl metrics** | | |
|---|---|---|---|---|---|---|---|---|
| | **MPEG D1 PSNR** | **$PSNR_{96,max}$** | **Gain** | **$PSNR_{98,min}$** | **Gain** | **MPEG D2 PSNR** | **$PSNR_{96,min}$** | **Gain** |
| **PCL** | 87.0 | 87.1 | **0.1** | 89.0 | **2.0** | 90.2 | 89.0 | -1.2 |
| **G-PCC** | 86.9 | 90.3 | **3.4** | 86.5 | -0.4 | 89.1 | 89.2 | **0.1** |
| **V-PCC** | 53.1 | 53.5 | **0.4** | 69.5 | **16.4** | 51.4 | 76.6 | **25.2** |
| **All codecs** | 67.3 | 76.3 | **9.0** | 77.6 | **10.3** | 68.9 | 75.8 | **6.9** |
| **Average** | 73.5 | 76.8 | **3.2** | 80.7 | **7.1** | 74.9 | 82.7 | **7.8** |

Since it is not the best solution to use different quality metrics for different point cloud codecs, unless variations of a specific codec are being compared, the correlation performance results have been analyzed to identify the generalized Hausdorff distance PSNR-based quality metric which is globally reliable. Although not the optimum for all point cloud codecs, the Po2Po generalized Hausdorff $PSNR_{96,max}$ shows better correlation compared to the MPEG D1-PSNR metric as shown in Table 6.4 for all codecs individually and also altogether. However, there are other generalized Hausdorff-based PSNRs that, while not always outperforming the MPEG metrics, show very large

correlation gains for the codec typically exhibiting lower objective-subjective correlations when using the MPEG metrics, this means V-PCC; however, a small correlation loss may be observed for one of the remaining codecs. As shown in Table 6.4, this is the case for Po2Pl $PSNR_{96,min}$ and Po2Po $PSNR_{98,min}$, which show rather large correlation gains (up to 25.2% and 16.4%, respectively) for V-PCC, the codec where correlation improvements are most needed. In summary, the Po2Pl $PSNR_{96,min}$ is a good metric to outperform the MPEG metrics for the emerging point cloud geometry coding solutions.

## 6.4 Final Remarks

This chapter shows that the proposed generalized Hausdorff-based PSNR quality metric outperforms the MPEG D1-PSNR and D2-PSNR point cloud geometry quality metrics for all considered codecs individually, and also altogether, if the appropriate rank (96%-99%) is selected. In this case, better objective-subjective correlation than the MPEG quality metrics is achieved for Po2Pl $PSNR_{96,min}$, thus indicating that these metrics may be used with advantage for future point cloud geometry quality assessment.

The work in this chapter led to a conference publication, notably:

- **A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "A generalized Hausdorff distance based quality metric for point cloud geometry," *IEEE International Conference on Quality of Multimedia Experience* (QoMEX), Athlone, Ireland, May 2020.

# Chapter 7

# Resolution-adaptive PSNR-based Quality Metric for Point Cloud Geometry

## 7.1 Context and Objectives

In Chapter 4, the subjective-objective correlation study shows that MPEG D1-PSNR and D2-PSNR quality metrics do not perform well to assess the quality of decoded point clouds, when point clouds with different precision are available and also when they are encoded with different (non-normative) codecs, such as PCL. In Chapter 5, the correlation performance of these quality metrics was evaluated in assessing the quality of decoded point clouds for the new point cloud coding standards, i.e., MPEG G-PCC and MPEG V-PCC. The experimental results show that they are also not performing well for these emerging standard codecs, especially for V-PCC.

D1-PNSR and D2-PSNR have shown low correlation performance in the literature, notably after the advent of the new point cloud geometry codecs, which tend to produce a very different number of decoded points in comparison to the number of reference points for a specific point cloud. Due to the way point clouds are acquired and pre-processed (before coding), their *intrinsic resolution*, a measure of the distance between points in the 3D space, plays an important role on the final perceived quality, not only to mitigate or highlight coding artefacts but also to measure the raw point cloud quality (i.e., after acquisition). The intrinsic characteristics of a point cloud that are going to be used in this chapter are as follows:

- **Precision:** regards the bit-depth of the geometry information in a voxelized point cloud; coordinates in a voxelized point cloud are limited between *0* and $2^{Precision}$ -1.
- **Intrinsic resolution:** measure of distance between points in a point cloud in 3D space; intrinsic resolution depends on precision.

- **Density:** number of points for some defined volume; it is a function of precision and intrinsic resolution.

These intrinsic characteristics have an impact on the perceived quality of point cloud after rendering. The distance between points is scaled after rendering to a target 2D size and resolution, and how they look depends on these intrinsic resolutions. Figure 7.1 shows *Egyptian Mask* with 12 and 20-bit precision, alongside with *Frog* with 12-bit precision. The bounding box of these point clouds are all scaled to the same image size. The maximum coordinate for 20-bit point clouds is $2^{(20-12)} = 265$ times larger than for 12-bit point clouds; while 10-bit and 12-bit *Egyptian Mask* look exactly similar because they have the same number of points, Frog seems denser because it has almost four times more points.

**Figure 7.1:** 12-bit *Egyptian Mask*, 20-bit *Egyptian Mask* and 12-bit *Frog*, their bounding boxes are scaled to the same size and rendered with the same 2D image size (from left to right).

On the other hand, distances in 3D are different from distances after projection on a plane (regardless of the rendering parameters and image size). Planar distance of a point in a voxelized point cloud to all its neighbors is equal but their 3D distance might be different based on the value of depth in a specific view. In this context, distance between point after rendering on a 2D display is a more precise measure for the intrinsic resolution of the 2D rendered point cloud. In this context, rendering resolution is defined as:

- **Rendering resolution:** measure of planar distance between points in a point cloud.

In this context, the main objectives of this chapter are:

- **Propose and evaluate geometry PSNR-based quality metrics that exploit the *intrinsic characteristics* of a point cloud**. In this case, the intrinsic resolution and precision are considered the most important intrinsic characteristics, influencing the final point cloud quality. For this purpose, the inhomogeneous distribution of points on the surface of the point cloud is considered to estimate the intrinsic resolution of a point cloud.
- **Propose and evaluate a PSNR-based metric exploiting the intrinsic or rendering resolution of a point cloud.** In this case, the intrinsic resolution is considered before and after

projection, leading to the so-called rendering resolution. This allows to significantly increase the quality metric performance, i.e., to obtain higher correlation with subjective scores.

The objective quality metrics to be proposed and evaluated in this chapter keep the simplicity of the state-of-the-art Po2Po and Po2Pl metrics and are based on the popular PSNR metrics currently used in the MPEG to evaluate the performance of the MPEG G-PCC and V-PCC codecs and new coding techniques.

The rest of this chapter is organized as follows. Section 7.2 describes and discusses the state-of-the-art on PSNR-based point cloud geometry quality metrics. Sections 7.3 presents the proposed quality metrics. Experimental results are presented and analysed in Section 7.4 and Section 7.5 concludes the chapter.

## 7.2 PSNR-based Geometry Quality Metrics

The state-of-the-art PSNR-based point cloud geometry quality metrics are presented in Chapter 2 (Section 2.6.3) as currently defined and used by the MPEG group [104]; they assess the geometry quality for compressed point clouds, independently of the codec used, target quality and rendering solution. The MPEG PSNR-based quality metrics are the basis for the new point cloud geometry quality metrics to be proposed in this chapter.

In the MPEG PSNR-based geometry quality metrics, the PSNR is obtained from a normalization factor and the mean squared error (MSE), as defined in equation (7.1), which is computed in two directions: from the decoded to the reference point cloud as well as in the opposite direction. The PSNRs for the two directions are then combined to obtain a single symmetric PSNR value using the maximum as the pooling function, as defined in equation (7.2).

$$\mathrm{PSNR}_{AB} = 10\log_{10}\left(\frac{p_s^2}{d_{AB}^{MSE}}\right) \tag{7.1}$$

$$\mathrm{PSNR} = \max\left(\mathrm{PSNR}_{AB}, \mathrm{PSNR}_{BA}\right) \tag{7.2}$$

In equation (7.1), $p_s$ is the signal peak and $d_{AB}^{MSE}$ is the mean squared error (i.e., MSE) of the distance, e.g., Euclidean distance, between all points in point cloud *A* and their corresponding nearest neighbor point in point cloud *B*. Point clouds *A* and *B* can be associated to reference and decoded point clouds, respectively. For more information on how MSE and MSE-PSNR are computed, please see Chapter 2.

As explained in Chapter 2, MPEG defines two PSNR-based metrics, the so-called D1-PSNR and D2-PSNR, that only differ on the type of distances used to compute the MSE. For D1-PSNR, MSE is computed for the Po2Po distance between each point in point cloud *A* and its nearest neighbor in point cloud *B*. For D2-PSNR, MSE is still computed for the distance between each point to its nearest neighbor, but this distance is now measured as the projection of the Po2Po error vector along the normal vector of the underlying surface at each point, which is known as Po2Pl distance.

Thus, D1-PSNR and D2-PSNR may be obtained using Po2Po and Po2Pl MSE (D1 and D2), in equation (7.1). As far as the signal peak $p_s$ in equation (7.1) is concerned, it typically corresponds to the largest diagonal (LD) distance for the point cloud bounding box for non-voxelized data [116]:

$$LD = \left\| (x_{max}, y_{max}, z_{max}) - (x_{min}, y_{min}, z_{min}) \right\|_2 \tag{7.3}$$

where $x$, $y$ and $z$ are the three geometry position coordinates. When LD of the point cloud bounding box is used as peak in the PSNR computation, the point coordinates are normalized to the range [0,1], what is equivalent to having both reference and decoded point clouds fitted to a unit size bounding box. However, since two point clouds may have different sizes, they can be scaled differently; this scenario may occur when the decoded point cloud has errors with a large magnitude, which may lead to a bounding box of different size.

If point clouds have been voxelized, the (point) coordinates lie on a regular 3D (integer) grid with some fixed, predefined precision, i.e., the coordinates bit-depth or point cloud precision. Thus, point coordinates are bounded between zero and a constant integer related to the point cloud precision. For voxelized point clouds, the peak for each coordinate represented with $b$ bit-depth precision is:

$$p_c = 2^b - 1 \tag{7.4}$$

By applying the peak for each coordinate in equation (7.3), the signal peak value corresponds to:

$$p_s = \sqrt{3} p_c \tag{7.5}$$

By using equation (7.5) in equation (7.1), point clouds are scaled to the same precision and the PSNR for the distance in question is computed as follows:

$$PSNR_{AB} = 10 \log_{10} \left( \frac{3 p_c^2}{d_{AB}^{MSE}} \right) \tag{7.6}$$

Note that increasing the point cloud precision makes distances between points larger in the new scale; for example, a point cloud with 11-bit precision has twice larger distances between points compared to the same point cloud with 10-bit precision. The normalization of errors proposed in equation (7.6) makes the comparison of data in different precisions possible by compensating for that difference. However, none of the PSNR-based quality metrics that are currently used, account for differences in:

1. The point cloud intrinsic resolution, since the perception of details depends on the sampling frequency.
2. Viewing conditions, namely the important rendering process that influences the final point cloud quality and can mitigate or highlight some coding distortions [117].

In practice, independently of the coding artifacts, the subjective evaluation of a given point cloud varies when these factors vary.

## 7.3 Proposed Point Cloud Geometry Quality Metrics

In this section, the proposed point cloud quality metrics for geometry are presented and explained. First, in Section 7.3.1, the intrinsic resolution of a point cloud is used to improve the objective-subjective correlation of the conventional, MPEG PSNR-based metrics, leading to the I-PSNR metrics. Then, in Section 7.3.2, the intrinsic resolution of a point cloud after rendering, so-called rendering resolution, is considered and used in the design of the resolution adaptive RA-PSNR metric.

### 7.3.1 Intrinsic Resolution PSNR-based Quality Metrics

In this section, besides the usual error to account for coding distortions, the point cloud intrinsic characteristics are also exploited to design improved point cloud geometry quality metrics.

In [118], the intrinsic resolution is used to normalize the geometry errors in the PSNR computation, as described in the previous section. The intrinsic resolution is difficult to measure since a point cloud is unstructured and the neighborhood of a point is more complex to define than for a pixel in a 2D image. However, the point cloud intrinsic resolution concept is analogous to the spatial resolution of a 2D image and can be estimated from the distance of points to their neighbors [119]. This distance may not be uniform throughout the entire point cloud and requires the design of suitable estimators. Simple methods to estimate the point cloud intrinsic resolution have been already proposed in the past. For example, [118] suggests using the maximum nearest neighbor ($MNN$) distance over all points in the point cloud. If $d_i$ is the distance from point $i$ to its nearest neighbor in the same point cloud $O$, then the intrinsic resolution can be estimated according to:

$$MNN = \max_{\forall i \in O} d_i \tag{7.7}$$

However, $MNN$ is very sensitive to holes and locally sparse areas, even if they are very small comparing to the point cloud size.

In this context, the following two intrinsic resolution estimators are proposed in this chapter to overcome this problem. First, the average nearest neighbor ($ANN$) is proposed, which attempts to solve the problem of localized sparse areas:

$$ANN = \sqrt{\frac{1}{N_O} \sum_{i \in O} d_i}^{\,2} \tag{7.8}$$

where $N_O$ is the number of points in the reference point cloud $O$.

Second, $ANN$ can be generalized if the average of $k > 1$ nearest neighbors ($ANN_k$) is used instead of only the closest nearest neighbor used in equation (7.2), which may not be very reliable due to acquisition noise. $ANN_k$ is defined as:

$$ANN_k = \sqrt{\frac{1}{N_0}\sum_{i\in O}\left(\frac{1}{k}\sum_{j=1}^{k} d^2\right)_{ij}} \quad (7.9)$$

In this context, the so-called Intrinsic resolution PSNR-based (I-PSNR) metric can be defined by using the intrinsic resolution estimators proposed above in equations (7.8) - (7.9) as normalizers to convert the mean squared errors to PSNR values. This means that the peak parameter, $p_s$, in equation (7.1) will correspond to one of the intrinsic resolution estimators computed over the reference point cloud using equations (7.7) - (7.9), instead of the largest diagonal distance of the bounding box (7.3) or the precision established in the D1-PSNR and D2-PSNR metrics (7.5). Thus, I-PSNR is defined as:

$$\mathrm{PSNR}_{AB}(ANN_k) = 10\log_{10}\left(\frac{\frac{1}{N_0}\sum_{i\in O}\left(\frac{1}{k}\sum_{j=1}^{k} d_{ij}^2\right)}{d_{AB}^{MSE}}\right) \quad (7.10)$$

In this way, point clouds with different distances between points (i.e., sparser, or denser) will be normalized accordingly, i.e., all MSE errors will be scaled according to the estimated intrinsic resolution. However, all these metrics still do not consider the rendering process that needs to be applied after point cloud decoding for visualization purposes.

### 7.3.2 Resolution-adaptive PSNR-based Quality Metrics

Nowadays, point clouds are still typically rendered as images or videos from one or more viewpoints to be shown on a 2D (or 3D) display, this means considering that some virtual observer is at some location in the 3D world with some virtual camera setup that determines which parts of the point cloud are shown to the user.

In this process, the distances between the points in a point cloud are scaled with respect to this viewing box and display resolution. The viewing box is defined by the virtual camera position, orientation, and characteristics (e.g., field of view) and the 3D to 2D projection (often a perspective projection is used). Since point clouds are always evaluated by the users after rendering, the final perceived quality does not only depend on the point cloud errors introduced by some processing step (in this case, coding) but also by the rendering process. In this section, a novel point cloud geometry quality metric is proposed based on the idea of estimating the intrinsic resolution after rendering, i.e., considering the distance of each point to its nearest neighbors on the image plane observable by the user. This is hereafter referred to as *rendering resolution* and a procedure to estimate it is proposed in the next section.

#### *7.3.2.1 Rendering Resolution Estimation*

The rendering resolution may vary for different parts of the point cloud due to several factors, e.g., orientation of the point cloud surfaces in the 3D world relative to the observer viewing location.

Since the viewing location and other rendering parameters, e.g., distance from the point cloud to the image plane, are not known when some geometry quality metric is computed, it is assumed that the point cloud is viewed from all possible directions at a fixed distance. This is a usual scenario in many applications (e.g., cultural heritage); however, further optimizations are possible such as using multiple distances (or scales) as other 2D quality metrics such as MS-SSIM do [120].

Consider that a unit normal vector $\vec{n}_i$ is available for every point $a_i$ in the point cloud. Normal vectors, which are already used in Po2Pl point cloud quality metrics, can be quickly estimated using some state-of-the-art method. The rendering resolution estimation process is illustrated in Figure 7.2 and proceeds as follows:

1. Define a local neighborhood $\gamma$ around point $a_i$, including the closest (nearest neighbors) $k$ points (represented by the six red dots in Figure 7.2).
2. Define a local plane, tangent at point $a_i$. This plane is perpendicular to the normal vector $\vec{n}_i$ and represents the point cloud surface at this point. This plane can also be seen as the image plane for the rendering process if the viewing direction (usually defined as a vector) is symmetric to $\vec{n}_i$ and the observer location lies along the viewing direction.
3. Project all points in the $\gamma$ neighborhood, represented by $\vec{d}_{ij}$ vectors in Figure 7.2, to the local plane tangent at point $a_i$ according to equation (7.11); in this equation, $_{ij}\vec{d}$ stands for the distance vector between point $a_i$ and its $j^{\text{th}}$ nearest neighbor. This step results in a vector for each projected point (represented by the red dashed arrows in Figure 7.2), called *planar distance vector* $\vec{\vec{P}}\vec{\vec{D}}_{ij}$, whose origin is $a_i$.

$$\vec{\vec{P}}\vec{\vec{D}}_{ij} = Proj_{plane}(\vec{d}_{ij}) = \vec{d}_{ij} - Proj_{\vec{n}_i}(\vec{d}_{ij}) \tag{7.11}$$

4. Estimate the rendering resolution as the average (planar) distance between point $a_i$ and their $k$ local neighbors on the tangent plane. This is performed for all points $a_i$ to obtain a global estimation for the entire point cloud; in this case, averaging over all points resulted in a better estimation. Thus, the rendering resolution corresponds to the average over the entire reference point cloud of the average (planar) distance (APD) between point $a_i$ and the $k$ nearest neighbors in $\gamma$, $APD_k$, as follows:

$$APD_k = \frac{1}{N}\sum_{i \in O}\left(\frac{1}{k}\sum_{j=1}^{k}\left\|\vec{\vec{P}}\vec{\vec{D}}_{ij}\right\|_2\right)^2 \tag{7.12}$$

Since $APD_k$ includes the distance between points considering that some projection was performed in the rendering process, it is expected that it better reflects the intrinsic perceived point cloud quality when used in a geometry quality metric. The proposed resolution adaptive PSNR (RA-PSNR) metric exploits this factor and is described in the next section.

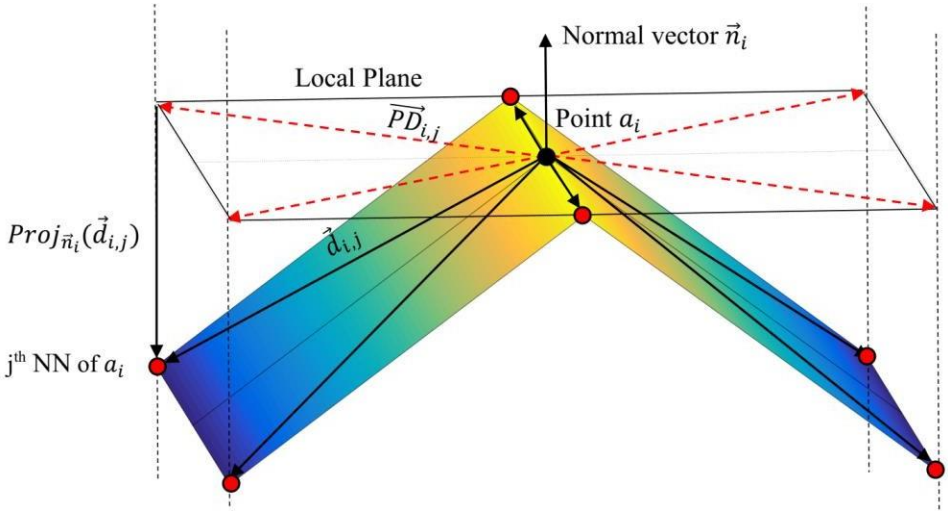



**Figure 7.2:** Illustration of the proposed rendering resolution estimation process.

### *7.3.2.2 RA-PSNR Point Cloud Quality Metric Design*

The proposed RA-PSNR point cloud geometry objective quality metric aims to assess the point cloud geometry quality more reliably by exploiting the rendering resolution as well as the precision used for the point cloud geometry coordinates. Ideally, the objective quality metric should compensate for any difference between point clouds in terms of these factors (e.g., precision), notably to avoid any undue influence on the estimated final perceived quality. For example, if the point cloud precision increases one bit without adding any new points, all the distances between points, used to calculate the intrinsic resolution, are now twice larger but the point cloud typically has a very similar perceived quality. To compensate for those factors, a density coefficient $D$ is defined as:

$$D = \frac{p_c}{r_{pc}} \tag{7.13}$$

where $p_c$ is the coordinate peak (related to the precision) calculated with equation (7.4), i.e., the maximum possible distance, and $r_{pc}$ is the point cloud resolution, which may be the intrinsic or rendering resolution. The proposed density coefficient corresponds to a normalization of the intrinsic or rendering resolution (which is typically defined as the inverse of density) according to the point cloud precision.

Using the state-of-the-art D1-PSNR and D2-PSNR metrics [104] defined by equation (7.6), all the errors (which represent a distance between points in the reference and degraded point clouds) are normalized according to the precision. Therefore, the intrinsic or rendering resolution of the point clouds proposed before in Section 7.3.1 and 7.3.2.1, which are also distances between points should also be normalized accordingly.

The proposed resolution adaptive quality metric, RA-PSNR, is defined in equation (7.14) and corresponds to using the density coefficient $D$ to further scale the MSE, taking into account the intrinsic point cloud characteristics and the rendering process in the objective quality assessment.

$$RA\text{-}PSNR_{AB} = 10\log_{10}\frac{3p_c^2}{Dd_{AB}^{MSE}} \tag{7.14}$$

By applying equation (7.13) in equation (7.14), RA-PSNR comes as:

$$\text{RA-PSNR}_{AB} = 10\log_{10}\frac{3r_{pc}p_c}{d_{AB}^{MSE}} \tag{7.15}$$

The parameter $r_{pc}$ can be the intrinsic resolution obtained from equations (7.7) to (7.9) or the rendering resolution obtained from equation (7.12), which are computed for the reference point cloud.

## 7.4 Performance Evaluation

This section evaluates the performance of the proposed point cloud quality metrics using point clouds coded with different codecs, notably the MPEG standard codecs.

### 7.4.1 Subjective Evaluation Dataset

In these experiments, the MOS obtained in a previous subjective test with DSIS methodology described in Chapter 5 are used. The MOS as well as the reference and decoded point clouds, referred as *IST Rendering Point Cloud Dataset*, are publicly available in [105]. The used point cloud dataset includes six point clouds from the MPEG repository [104], which have been coded with three rates/qualities using three rather different point cloud codecs: i) PCL octree-based codec; ii) MPEG G-PCC standard; and iii) MPEG V-PCC standard.

The quality scores were obtained at a test session where the decoded point clouds were rendered with a popular point-based rendering approach with uniform color and shading (*RPoint*). Because there are no coloring and interpolation processes involved in this rendering, geometry coding artefacts are less masked.

To evaluate the objective-subjective correlation performance for the proposed point cloud geometry quality metrics, a non-linear regression has been used to map the computed objective quality scores (PSNR) into the MOS scale. Thus, based on Recommendation ITU-T P.1401 [33], a monotonic cubic function was used to fit the objective scores to MOS and obtain predicted MOS values as:

$$\text{MOS}_p = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^4 \tag{7.16}$$

where $x$ are the objective metric scores and $\beta_1, \ldots, \beta_4$ the regression model parameters. Then, the PLCC and SROCC are used to assess the objective-subjective correlation, similarly to the previous chapter. The following point cloud quality metrics were considered:

1. **MPEG PSNR-based metrics:** These benchmark metrics represent the state-of-the-art and were described in Section 7.2. Two variants are defined, notably using the precision ($P$) and the largest diagonal ($LD$) distance of the bounding box as the signal peak.

2. **Proposed I-PSNR metrics:** PSNR-based metrics where, $ANN$ and $ANN_k$, the two proposed intrinsic resolution estimators, described in Section 7.3.1, are used, alongside with $MNN$ which was used before in the literature. The intrinsic resolution is used as the signal peak for the MPEG PSNR-based metrics described in Section 7.2.

3. **Proposed RA-PSNR metrics:** PSNR-based metrics where the proposed intrinsic and rendering resolution estimators $ANN$, $ANN_k$ and $APD_k$ (variants) described in Section 7.3.1 and 7.3.2.1 are used. The RA-PSNR metric corresponds to equation (7.15) as presented in Section 7.3.2.2.

For the experimental results shown here, the $k$ parameter for $ANN_k$ and $APD_k$ was set to 10; this was experimentally found as the value reaching the largest PLCC with MOS. For all quality metrics, the performance of all proposed normalization factors, i.e., $P, LD, MNN, ANN, ANN_k$ and $APD_k$, is shown for the Po2Po and Po2Pl distances.

### 7.4.2 Experimental Results

Table 1 shows the objective quality metrics performance for each point cloud codec and considering all codecs' data. Based on the correlation performance results presented in this table, the following conclusions can be made:

- **Proposed point cloud quality metrics vs benchmarks:** The benchmark PSNR-based metrics are outperformed by the proposed RA-PSNR and I-PSNR metrics for all codecs individually and all codecs data together ('All' column). I-PSNR metrics can achieve higher performance than the D1-PSNR and D2-PSNR metrics for PCL and V-PCC while RA-PSNR is consistently better for all cases. The best overall correlation performance is achieved for RA-PSNR with the rendering resolution $APD_k$ variant, which was expected since it considers the rendering process required for point cloud visualization. Moreover, for the 'All' case, where different types of coding artefacts and point cloud characteristics are accounted, for Po2Po metric, 6.7% gain was achieved for RA-PSNR with the $APD_k$ variant in both PLCC and SROCC comparing to MPEG D1-PSNR. For Po2Pl metric 5.3% gain in PLCC and 8.2% gain in SROCC were achieved comparing to MPEG D2-PSNR.
- **Quality metrics correlation performance for PCL and G-PCC codecs:** For PCL and G-PCC decoded data, the RA-PSNR proposed metric with the $APD_k$ variant can reach a very high correlation performance (95.2% and 94.0%, respectively for PCL and G-PCC), since it considers both the precision and rendering resolution, i.e., the point cloud density observed by the users.

- **Quality metrics correlation performance for V-PCC codec:** For V-PCC decoded data, the I-PSNR $ANN_k$ variant performs slightly better than RA-PSNR $APD_k$, because all point clouds considered are coded in 10-bit with V-PCC and the intrinsic resolution is already a good estimator of the quality after rendering. Normalizing based on precision acts as a distractor and decreases the impact of intrinsic resolution. In a general way, the benchmark PSNR-based metrics are rather poor for V-PCC data and both the proposed I-PSNR and RA-PSNR metrics are able to increase the correlation performance significantly, with 28% and 26% gains respectively, compared to the Po2Pl PSNR metrics.

**Table 7.1:** Correlation performance for the proposed and benchmark quality metrics. First two rows correspond to D1-PSNR and D2-PSNR.

| Metric | Variant | Type | PCL | | G-PCC | | V-PCC | | All | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | PLCC | SROCC | PLCC | SROCC | PLCC | SROCC | PLCC | SROCC |
| PSNR | P | *Po2Po* | 87.0 | 73.9 | 86.9 | 87.4 | 53.1 | 62.0 | 67.3 | 64.7 |
| | | *Po2Pl* | 89.6 | 80.9 | 83.4 | 85.6 | 51.4 | 49.6 | 70.3 | 65.6 |
| | LD | *Po2Po* | 83.3 | 82.3 | 86.0 | 89.3 | 48.9 | 54.1 | 70.4 | 68.6 |
| | | *Po2Pl* | 86.7 | 85.9 | 75.6 | 71.9 | 59.9 | 58.9 | 71.4 | 67.2 |
| I-PSNR | MNN | *Po2Po* | 68.7 | 65.0 | 40.8 | 33.3 | 45.6 | 14.3 | 49.7 | 42.2 |
| | | *Po2Pl* | 69.6 | 66.7 | 44.1 | 39.7 | 49.8 | 25.4 | 52.1 | 43.4 |
| | ANN | *Po2Po* | 92.2 | 89.0 | 79.3 | 76.3 | 66.6 | 61.0 | 64.7 | 52.5 |
| | | *Po2Pl* | 92.3 | 84.5 | 86.6 | 79.0 | 76.2 | 69.1 | 66.4 | 55.6 |
| | $ANN_k$ | *Po2Po* | 88.8 | 87.0 | 78.1 | 68.4 | 70.8 | 62.9 | 66.5 | 59.2 |
| | | *Po2Pl* | 90.8 | 87.5 | 75.3 | 68.2 | **79.6** | **74.0** | 67.4 | 62.6 |
| RA-PSNR | ANN | *Po2Po* | 92.8 | 86.8 | 84.9 | 82.5 | 49.2 | 45.6 | 68.9 | 64.0 |
| | | *Po2Pl* | 93.4 | 86.7 | 88.5 | 85.8 | 67.9 | 63.0 | 71.0 | 67.1 |
| | $ANN_k$ | *Po2Po* | 94.1 | 86.2 | 93.6 | 94.6 | 68.5 | 62.5 | 74.1 | 71.1 |
| | | *Po2Pl* | 95.1 | 91.3 | 94.0 | 94.0 | 77.1 | 73.7 | 74.9 | 72.5 |
| | $APD_k$ | *Po2Po* | 94.1 | 85.2 | 93.7 | **94.8** | 59.9 | 59.7 | 74.0 | 71.4 |
| | | *Po2Pl* | **95.2** | **91.7** | **94.0** | 94.0 | 77.9 | 72.3 | **75.6** | **73.8** |

## 7.5 Final Remarks

D1-PSNR and D2-PSNR point cloud geometry quality assessment metrics generally perform poorly since they only consider point cloud precision and the impact of the intrinsic point cloud characteristics and the rendering process on the final perceived point cloud quality are not considered. In this chapter, the popular PSNR-based metrics are improved by including a normalization factor that accounts for changes in the intrinsic point cloud resolution after

rendering, as well as the point cloud precision. Experimental results show that the proposed quality metrics outperform the MPEG PSNR-based quality metrics by a significant margin. As future work, more characteristics of the rendering process could be included, notably the viewing position and the point cloud distance to the viewing plane.

This work led to a conference publication, notably:

**A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Improving PSNR-based Quality Metrics Performance for Point Cloud Geometry," *IEEE International Conference on Image Processing* (ICIP), Abu Dhabi, UAE, October 2020.

# Chapter 8

# Mahalanobis-based Point-to-Distribution Quality Metric for Point Cloud Geometry

## 8.1 Context and Objectives

At the time this work was developed, the state-of-the-art point cloud geometry quality metrics were the MPEG PSNR-based metrics [104], notably relying on distances computed from point-to-point correspondences between the degraded and reference point clouds. However, correlation performance evaluation studies [117] [121] have shown that these metrics underperform for data coded with the MPEG point cloud codecs. A key characteristic of these codecs is that the number of decoded points is often much larger than the number of points in the reference point cloud. When additional decoded points are obtained, and they follow the underlying point cloud surface, higher perceptual quality typically results. However, the MPEG (and others) point cloud quality metrics are not able to properly account for this effect, since distances (seen as errors) are computed based on pointwise correspondences. In fact, the additional decoded points typically have their closest neighbor in the reference point cloud far away, which leads to large errors being accounted in the final quality metric value.

Since point clouds have their surface reconstructed during the rendering process before visualization, the objective quality metrics should properly model the underlying point cloud surface to achieve better objective-subjective correlation. In fact, the D2-PSNR metric follows this approach by parametrically modelling the underlying surface with a plane tangent at each point, defined by the point's normal. However, this is still a rather limited approach considering the unstructured point cloud nature, especially for regions that are not planar and for points far away from the plane. In this context, the main objectives of this chapter are:

- **Propose a better type of correspondence** between point clouds, notably when the decoded point clouds have many more points than reference point clouds.
- **Propose a point cloud geometry quality metric** that considers the underlying surface of points more efficiently than available point-based quality metrics.

In this chapter, a novel so-called *Mahalanobis-based point-to-distribution* quality metric is proposed, which adopts a new type of correspondence between two point clouds, namely between a point in one point cloud and the distribution associated to a limited set of points (from a small region) in another point cloud. The idea underpinning this novel type of point cloud geometry quality metric is to statistically characterize the point cloud surface, notably through the covariance of points within some local region. Moreover, this novel quality metric is not unduly influenced by the number of decoded points (compared to reference points), but rather by a statistical characterization of the points location. In addition, it is still reliable when the degraded and reference point clouds have different characteristics in terms of precision, density, and structure.

The rest of this chapter is organized as follows. Section 8.2 describes the proposed quality metric and Section 8.3 presents the experimental results along with their analysis. Finally, Section 8.4 ends the chapter with conclusions and future work suggestions.

## 8.2 Proposed Point Cloud Geometry Quality Metric

The proposed class of point-to-distribution (P2D) quality metrics works by computing the distance between a point on the reference (or degraded) point cloud and a distribution of points on a small region of the degraded (or reference) point cloud. This novel proposed approach has the following characteristics:

1. **Scale invariant:** can evaluate point clouds with different intrinsic characteristics in terms of precision (i.e., coordinates bit-depth) and resolution (i.e., average distance between neighboring points).
2. **Correlation aware:** considers the correlation of points with their neighbors and weights distances in each coordinate accordingly. Points not following the distribution of the surface points will lead to larger distances.
3. **Normal computation-free:** does not require the estimation of normal vectors to model the underlying surface, which is an error-prone and computationally complex process.

These characteristics allow to achieve better objective-subjective correlation performance compared to the point-based metrics, either adopting a single point (no surface) or a tangent plane at a point, to characterize the point cloud.

This work adopts the *Mahalanobis* distance to measure the distance between a point and a distribution. This distance, which has already several applications in the areas of anomaly detection [122] and classification [123], is multivariate and exploits correlations between the point coordinates. Moreover, it is more suitable compared to the Euclidean distance (used in the Po2Po

and Po2Pl metrics) when there is some spatial correlation in the data, which is the case for a local region of points in a point cloud.

### 8.2.1 Directed Mahalanobis-based P2D Distance

Figure 8.1(a) shows the proposed *Mahalanobis*-based directed P2D distance architecture while Figure 8.1(b) shows the final symmetric distance and PSNR-based metric variants' architecture, respectively.

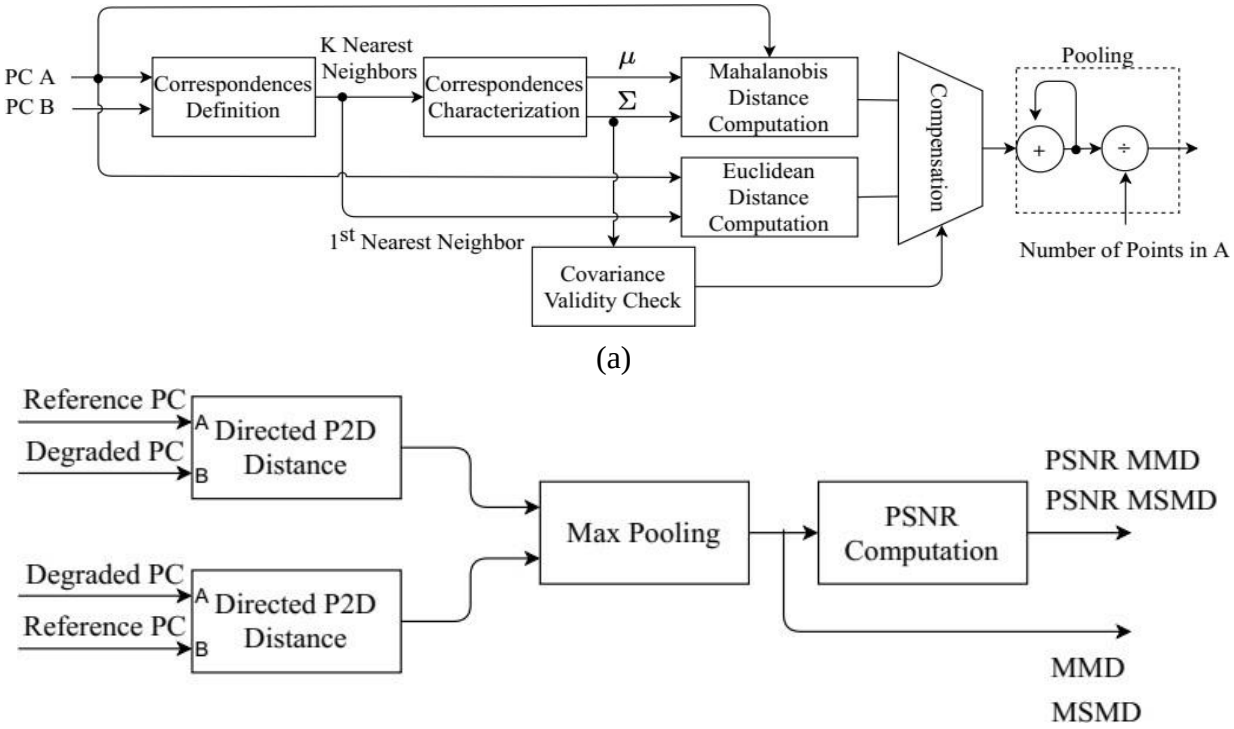



**Figure 8.1:** *Mahalanobis* based P2D quality metric architecture: a) directed (point cloud *A* to point cloud *B*) P2D distance; b) symmetric P2D distance and PSNR-based quality metric variants.

As Po2Po and Po2Pl metrics, this Mahalanobis-based distance is computed in both directions, i.e., reference to degraded (i.e., from point cloud *A* to *B*) and vice-versa. For point $\vec{a} = (x, y, z)^T$ in point cloud *A*, the directed P2D distance is computed as:

1 **Correspondences Definition:** Find the set of points $D_a$ corresponding to the $K$ nearest neighbors of $\vec{a}$ in point cloud *B*. This local set of points represents the support region for the metric in point cloud *B*. Naturally, the $K$ value is rather important to define the size of the local neighborhood; when $K$ is too small, spatial correlation between points may not be fully captured and, when *K* is too large, fine details may be missed.

2 **Correspondences Characterization:** Compute the mean $\vec{\mu}_{D_a}$ and covariance matrix $\Sigma_{D_a}$ for the $K$ neighboring points in point cloud *B*. The mean vector (size 3) contains the arithmetic average of the points coordinates, while the covariance matrix expresses the $(x, y, z)$ coordinates correlation.

$$\vec{\mu}_{D_a} = \frac{1}{K}\sum_{i=1}^{K}\vec{p}_i \tag{8.1}$$

$$\Sigma_{D_a} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix} \tag{8.2}$$

where $\vec{p}_i$ is the $i^{th}$ nearest neighbor of point $\vec{a}$, $K$ is the number of nearest neighbors and $\sigma_{PQ}$ is the covariance of variables $P$ and $Q$ with $P, Q \in \{x, y, z\}$. While the mean vector represents the centroid, the covariance matrix represents the data dispersion for the selected set of points in point cloud *B*.

3 **Covariance Validity Check**: Check if the covariance matrix is singular or ill-conditioned. While in the former case the *Mahalanobis* distance cannot be computed, in the latter case it may lead to very large *Mahalanobis* distances. Intuitively, this may occur when there is a linear dependence between the K-nearest neighbor points coordinates. The covariance matrix is singular when the determinant is equal to zero and ill-conditioned when the ratio of largest and smallest singular values in the singular value decomposition (condition number) is too large. Ill-conditioned covariance matrices can be found with the Cholesky decomposition or by thresholding on large condition numbers [124], which is the approach used here. Since in this case the covariance matrix is real and symmetric, the condition number is simply the ratio of the largest eigenvalue $\lambda_{MAX}$ to the smallest $\lambda_{MIN}$ eigenvalue of $\Sigma_{D_a}$and, thus, the covariance validity check comes as:

$$C > \epsilon, \quad \text{with } C = \lambda_{MAX}/\lambda_{MIN} \tag{8.3}$$

where $\epsilon$ is the validity check threshold; experimentally, it was found that $\epsilon = 10^6$ identifies well the ill-conditioned cases.

4 **Mahalanobis Distance Computation:** The *Mahalanobis* distance between point $\vec{a}$ and distribution $D_a$ is:

$$d_{aB}^{P2D} = \sqrt{(\vec{a} - \vec{\mu}_{D_a})^T \Sigma_{D_a}^{-1} (\vec{a} - \vec{\mu}_{D_a})} \tag{8.4}$$

i.e., the distance between $\vec{a}$ and $\vec{\mu}_{D_a} = (\mu_x, \mu_y, \mu_z)^T$, divided by the covariance $\Sigma_{D_a}$. Thus, if the $K$ nearest neighbor points are correlated in one or more coordinates (e.g., planar surface), the covariance is high, which will reduce the distance between $\vec{a}$ and $\vec{\mu}_{D_a}$. On the other hand, if points are not very correlated, the covariance is low, and this distance is not much reduced.

Moreover, if point $\vec{a}$ is located at the mean of the distribution, the *Mahalanobis* distance is zero and increases as $\vec{a}$moves away from the mean.

5 **Euclidean Distance Computation:** The Euclidean distance between point $\vec{a}$ and its nearest neighbor in point cloud *B* is computed. This is equivalent to the *Mahalanobis* distance with an identity covariance matrix.

6 **Compensation**: For the cases where the covariance matrix is ill-conditioned or singular (see Step 3), the Euclidean distance computed in Step 5 is used; otherwise, the *Mahalanobis* distance computed in Step 4 is used.

The process above is applied to all points in point cloud $A$ to obtain the *Mahalanobis* or Euclidean distance. Finally, all the computed distances are integrated by performing:

7 **Pooling**: Average pooling is used to aggregate the distances obtained for all points in point cloud $A$. Both the average of distances and squared distances may be used:

$$d_{A,B}^{\mathrm{MMD}} = \frac{1}{N_A} \sum_{a \in A} d_{a,B}^{P2D} \tag{8.5}$$

$$d_{A,B}^{\mathrm{MSMD}} = \frac{1}{N_A} \sum_{a \in A} \left(d_{a,B}^{P2D}\right)^2 \tag{8.6}$$

where MMD stands for mean *Mahalanobis* distance, MSMD for mean squared *Mahalanobis* distance and $N_A$ is the number of points in point cloud $A$.

### 8.2.2 Symmetric Mahalanobis-based P2D Distance

To obtain a symmetric distance, the directed distance described in the previous section is computed in the other direction, i.e., between every point $b$ in point cloud $B$ and the corresponding distribution in point cloud $A$. The symmetric distance is obtained as the maximum of the two directed distances:

$$\mathrm{MMD} = \max\left(d_{A,B}^{\mathrm{MMD}}, d_{B,A}^{\mathrm{MMD}}\right) \tag{8.7}$$

$$\mathrm{MSMD} = \max\left(d_{A,B}^{\mathrm{MSMD}}, d_{B,A}^{\mathrm{MSMD}}\right) \tag{8.8}$$

where MMD/MSMD are the final distances. Both distances can represent the error between two point clouds, often a reference and a decoded point clouds, and thus can be directly used as point cloud quality metrics although this is not necessarily the best solution.

### 8.2.3 PSNR-based Mahalanobis P2D Point Cloud Quality Metric

The *Mahalanobis* distance is a unit-less and scale-invariant distance that allows to account for the point cloud spatial correlation. Since large distances are avoided by checking the ill-conditioning of the covariance matrix, MMD and MSMD never exceed the maximum Euclidean distance. For voxelized data, all coordinates are bounded between 0 and $p_c = 2^{pr} - 1$ with $pr$ as the point cloud

precision and $p_c$ the peak coordinate value. Thus, to compute the PSNR-based point cloud quality metric from the distances aforementioned, the maximum diagonal Euclidean distance ($\sqrt{3}p_c$) is used as the peak for PSNR computation. The MMD and MSMD PSNR-based quality metrics come:

$$PSNR_{MMD} = 10\log_{10}(3p_c^2/\mathrm{MMD}^2) \tag{8.9}$$

$$PSNR_{\mathrm{MSMD}} = 10\log_{10}(3p_c^2/\mathrm{MSMD}) \tag{8.10}$$

The used point cloud precision corresponds to the maximum precision for all point cloud test materials (12-bit for the experimental results next). This fixed value assures that the same normalization is applied to all point clouds under evaluation, which allows to measure the correlation performance for point clouds with different precisions.

## 8.3 Performance Evaluation

In this section, the proposed quality metrics performance is assessed, notably in comparison with state-of-the-art metrics, using point clouds coded with different type of point cloud codecs.

### 8.3.1 Subjective Evaluation Dataset

To evaluate the proposed quality metrics defined by equations (8.7) - (8.10), the MOS and the reference and decoded point clouds from the *IST Rendering Point Cloud Dataset*, publicly available in [105], have been used. The dataset includes six point clouds from the MPEG repository [104], which have been coded with three rates/qualities using the following point cloud codecs: 1) PCL (octree-based); 2) MPEG G-PCC; and 3) MPEG V-PCC. All decoded point clouds were evaluated with a DSIS subjective assessment protocol. MOS for the point-based rendering session (RPoint, see Chapter 2) were used since geometry coding artefacts are more visible for this type of rendering, notably due to the absence of color attributes and mesh-based point cloud reconstruction.

A monotonic cubic function has been adopted to fit the objective metric values to the MOS and obtain the predicted MOS scores according to Recommendation ITU-T P.1401 [33]:

$$\mathrm{MOS_p} = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 \tag{8.3}$$

where $x$ are objective metric values and $\beta_1, \ldots, \beta_4$ are the regression model parameters. However, the results are also presented without the fitting function since overfitting mat happen considering the reduced amount of data. Figure 8.2 shows the scatter plots of different metrics versus MOS with the fitted curve, used to obtain predicted MOS values. As expected, objective scores are more correlated with the corresponding MOS, for the proposed MMD and MSMD PSNR point cloud quality metrics.

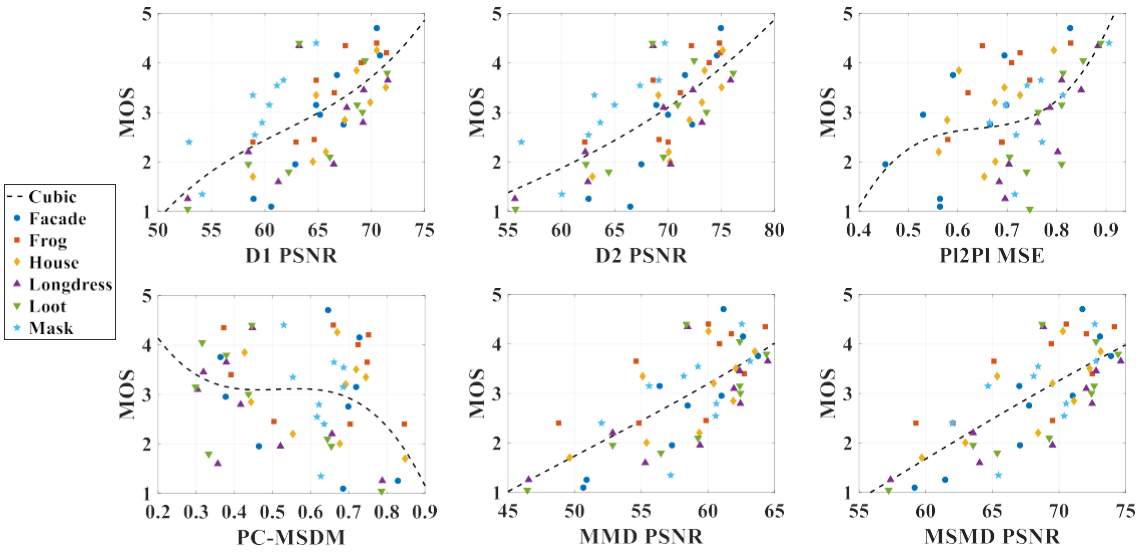



**Figure 8.2:** Subjective vs objective results

Similar to previous chapters, the objective-subjective correlation is assessed with the PLCC and SROCC. The proposed P2D *Mahalanobis-based* metrics are compared to the D1-PSNR and D2-PSNR quality metrics (both distance and PSNR variants) as well as to the Pl2Pl MSE (the best variant) [109] and the PC-MSDM [125] metrics. For the Pl2Pl quality metric, normal vectors were computed using the CloudCompare software [83] with default settings.

### 8.3.2 Neighborhood Size Selection

The correspondences definition step in the proposed *Mahalanobis-based* quality metrics requires the selection of an appropriate value for the $K$ parameter (number of nearest neighbors), since it has a major impact on the proposed metrics correlation performance.

As long as the decoded points are sampled from the same distribution, increasing $K$ should increase the quality metric reliability, and thus, larger $K$ values should be preferred. However, if $K$ is increased too much, then the nearest neighbors set will include less correlated points, not following the same distribution, which lowers the metric performance. Figure 8.3 shows the PLCC correlation for the MMD and MSMD PSNR metrics (with no fitting) as a function of the $K$ parameter. For all adopted point cloud codecs, increasing the neighborhood size increases the proposed metrics' performance up to the point where the correlation performance stays somewhat stable, slightly increasing or decreasing for some specific codecs.

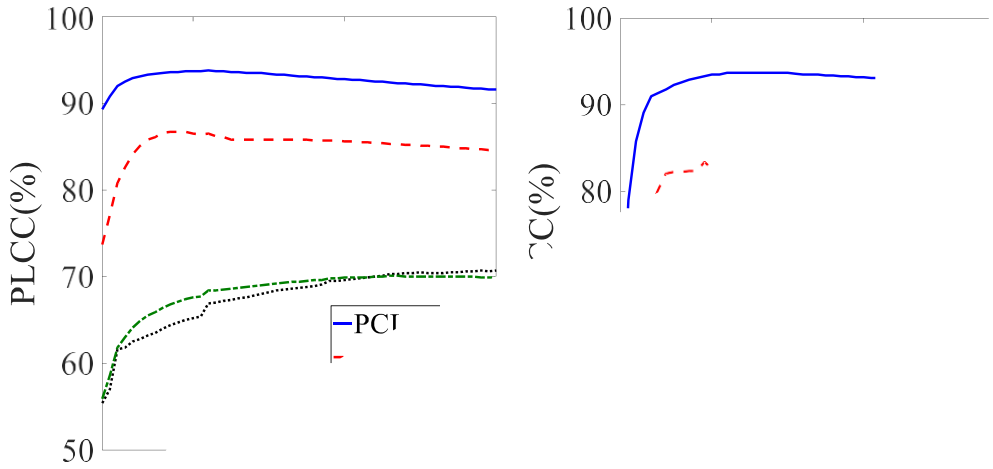



**Figure 8.3:** PLCC performance as a function of the nearest neighbor size (***K***).

For the proposed quality metrics, a fixed $K$ parameter is used to avoid overfitting. The $K$ parameter is selected based on the rank of their PLCC and SROCC correlation measures. More precisely, for each $K$, $PLCC_k$ and $SROCC_k$ are computed and the rank $R_k(c, m)$ determined for each coding configuration $c$, with $c$ = {PCL, G-PCC, V-PCC, All}, for both correlation measures $m$, i.e., PLCC and SROCC. Then, the selected rank $K^{opt}$ can be defined as $K^{opt} = \underset{k}{\operatorname{argmin}}\{\mathbb{E}(R_k(c, m))\}$ (lower ranks correspond to higher correlations), for every $c$ and $m$.

$K^{opt}$ is 31 for MMD and 38 for MMSD, both for the distance and PSNR-based variants.

Figure 8.4 shows the average percentage of points with an invalid covariance matrix using as reference the original point cloud and as degraded the decoded point cloud (and vice versa). This result follows the test conditions (codecs and bitrates) given in Section 8.3.1. As expected, the percentage of invalid covariance matrix decreases as $K$ increases and for the selected $K$ parameter values, the average percentage of ill-conditioned or singular covariance matrices is less than 5 percent.

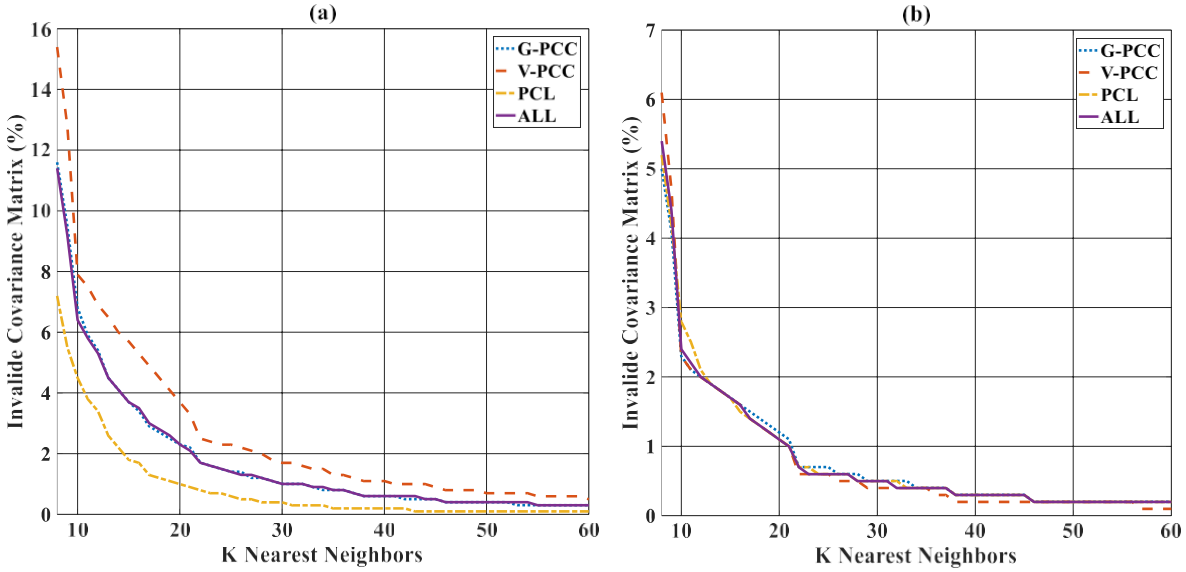



**Figure 8.4:** Percentage of points that their corresponding distribution has an invalid Mahalanobis distance for different values of nearest neighbor size in both directions: a) reference to degraded b) degraded to reference.

### 8.3.3 Experimental results

Table 8.1 shows the correlation performance for several quality metrics, for each individual point cloud codec and for 'All' codecs (i.e., all data jointly considered), before and after cubic fitting; The following conclusions may be taken:

- **No fitting vs cubic fitting:** Two different cases can be identified: 1) for the proposed quality metrics, non-fit correlations are closer to cubic fitting correlations, thus showing that these metrics can generalize better; 2) for the Po2Po and Po2Pl metrics, the difference is much larger, thus showing that their correlation performance is more sensitive to the cubic interpolation parameters fitting.
- **MMD vs MSMD:** PLCC and SROCC show that MMD metrics outperform MSMD metrics for each of the considered codecs individually, especially for the distance-based metrics; however, MSMD outperforms MMD when the data for 'All' codecs is considered together. Moreover, the MMD and MSMD PSNR-based variants have higher PLCC and, thus, express better the perceived quality in comparison to distance-based metrics, especially if no fitting is considered.
- **Proposed vs MPEG metrics:** Almost for all cases, the proposed distance and PSNR-based metrics have higher PLCC and SROCC than the MPEG D1 and D2 metrics. The performance gain is the largest for PCL and V-PCC, with PLCC gains up to 5.9% and 31.9%, respectively, and SROCC gains up to 13% and 22.8%, respectively, when cubic fitting is applied, considering the best proposed (in bold) and MPEG metrics (bold red).

- **V-PCC vs PCL vs G-PCC:** Larger performance gains are obtained for the V-PCC codec (up to 31.9% PLCC), compared to other codecs (5.9% for PCL and 2.5% for G-PCC). This is because V-PCC adds more additional decoded points that follow the underlying point cloud surface [117] in comparison to other point cloud codecs.

**Table 8.1:** PLCC and SROCC correlation performance for the proposed quality metrics in comparison to state-of-the-art metrics.

| | | PLCC | | | | | | | | SROCC | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Type** | **Quality Metric** | **PCL** | | **G-PCC** | | **V-PCC** | | **All codecs** | | **PCL** | | **G-PCC** | | **V-PCC** | | **All codecs** | |
| | | *No Fit* | *Cubic* | *No Fit* | *Cubic* | *No Fit* | *Cubic* | *No Fit* | *Cubic* | *No Fit* | *Cubic* | *No Fit* | *Cubic* | *No Fit* | *Cubic* | *No Fit* | *Cubic* |
| Po2Po | *D1 distance* | 29.3 | 83.7 | 50.8 | 52.5 | 6.9 | 23.8 | 33.6 | 52.1 | 74.3 | 72.2 | 65.4 | 64.9 | 32.3 | 37.8 | 62.4 | 62.2 |
| | *D1-PSNR* | 84.3 | 87.0 | **83.7** | **86.9** | 23.4 | **52.5** | 67.1 | 67.3 | 75.7 | 73.8 | **87.4** | **87.4** | 35.4 | 60.7 | 64.8 | 64.8 |
| Po2Pl | *D2 distance* | 29.8 | 83.6 | 47.6 | 48.8 | 26.2 | 30.0 | 29.6 | 47 | **81.7** | **81.0** | 63.3 | 63.7 | 26.1 | 30.8 | 59.1 | 58.8 |
| | *D2-PSNR* | **86.9** | **89.6** | 79.4 | 83.4 | **40.3** | 51.5 | **70.0** | **70.4** | **81.7** | 80.9 | 85.6 | 85.6 | **44.3** | **51.9** | **65.7** | **65.7** |
| Pl2Pl | *MSE [9]* | 71.8 | 72.5 | 48.6 | 55.0 | 35.6 | 52.7 | 45.1 | 51.0 | 73.4 | 73.4 | 40.1 | 40.1 | 36.8 | 58.6 | 45.0 | 45.0 |
| *PC-MSDM [11]* | | 68.5 | 69.9 | 0.7 | 68.4 | 29.7 | 80.9 | 24.6 | 31.6 | 65.5 | 65.5 | 28.9 | 74.2 | 45.7 | 77.4 | 18.2 | 16.4 |
| P2D | *MMD* | 89.6 | 93.9 | 80.8 | 88.6 | 60.6 | **84.4** | 66.5 | 70.0 | **92.7** | 92.7 | 85.3 | 85.6 | **75.9** | **83.5** | 65.7 | 65.7 |
| | *MMD PSNR* | **93.4** | 94.2 | **86.1** | 87.8 | **68.9** | 84.2 | 70.0 | 70.0 | **92.7** | 92.7 | 85.1 | 85.1 | **75.9** | 80.5 | 65.7 | 65.7 |
| | *MSMD* | 82.8 | **95.5** | 72.3 | **89.4** | 43.4 | 80.2 | 65.2 | 73.1 | **92.7** | **94.0** | 82.2 | 84.3 | 75.2 | 78.2 | **69.4** | 68.3 |
| | *MSMD PSNR* | **93.3** | 94.0 | 84.4 | 85.1 | 64.8 | 80.9 | **73.6** | **73.4** | **92.7** | 92.7 | 82.2 | 81.4 | 75.2 | 78.2 | **69.4** | **68.4** |

Finally, the MMD PSNR metric has the best correlation performance since it outperforms MSMD PSNR for all individual codecs (in PLCC and SROCC), although not for 'All' codecs case. This means that, for optimizing or monitoring the quality of a specific codec, the MMD PSNR is the best choice. Moreover, it outperforms Pl2Pl and the PC-MSDM metric, which were more emerging quality metrics at the time this work was developed.

## 8.4 Final Remarks

This paper proposes a novel type of point cloud quality metric where the distance is computed between a point and a set of points following some distribution. This allows to improve the objective-subjective correlation performance comparing to state-of-the-art point cloud geometry quality metrics, especially for point clouds with different characteristics in terms of precision, density, and structure. As future work, a point to distribution metric following the same approach but targeting point cloud attributes, notably color could be developed.

This work led to a journal publication, notably:

**A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, "Mahalanobis Based Point to Distribution Metric for Point Cloud Geometry Quality Evaluation," *IEEE Signal Processing Letters*, vol. 27, pp. 1350-1354, July 2020.

# Chapter 9

# Geometry Aligned Projection-based Quality Metric for Point Cloud

## 9.1 Context and Objectives

As described in the literature review of Chapter 2, a few works explored the idea of estimating the point cloud quality by projecting the point cloud into one or more 2D images, i.e., by converting a 3D point cloud into several 2D images [126] [127] [121] [128], and then applying a 2D quality metric. In the field of point cloud compression, this type of approach was successfully exploited to achieve efficient compression, as demonstrated by the MPEG V-PCC standard. In the context of point cloud quality assessment, these 2D images can be obtained by performing projection from different viewpoints, i.e., using different projection centers. Then, recent, and powerful 2D quality metrics can be used, to assess the entire point cloud quality. However, current *projection-based metrics* available in the literature are not showing better correlation performance than Po2Po quality metrics, where correspondences are established in the 3D space and errors in position or color are accounted.

One of the problems in projection-based metrics is that geometry distortions cause displacement errors between the projected reference and degraded images. However, 2D quality metrics do not handle well local displacement errors (or geometry errors) since comparisons are made pixel-wise (or region-wise). Typically, 2D objective quality metrics consider that these regions as high distortion when small or medium geometry errors are perceptually well tolerated when color is available (see Chapter 5). For example, small displacement errors are not perceived by humans but may lead to high distortions when conventional 2D quality metrics are used. Another issue arises when the number of points in reference and decoded point clouds is different. This often occurs, since some point cloud compression engines often use planar or other polygonal (e.g.,

triangular) approximations of the point cloud surface and thus, more points may be recreated at the decoder side when these surfaces are sampled. This means that the degraded 2D images may have, for some positions, pixels occupied when this pixel is not filled in the 2D image. In some past work [127] [126], these positions are ignored, or an occupied position is compared with a non-occupied position (usually filled with some background color). However, neither of these solutions are acceptable, since for some cases, these pixels are visually important (and can contribute to a higher quality) and the quality score cannot depend on the selected background color.

In this chapter, a novel projection-based point cloud quality metric is proposed, which addresses these limitations and achieve high performance in terms of objective-subjective correlation. The main ideas underpinning this quality metric is to compare the projected reference and degraded color images for two fixed levels of geometry, for the reference and decoded geometry. The two scores obtained for both geometry levels are fused together which means that the proposed approach implicitly considers geometry distortion as well as color distortions. By comparing images with the same geometry, any misalignment is avoided. Moreover, this novel quality metric can handle decoded point clouds with large or small number of points with respect to the reference, by performing padding in the 2D domain, this means using interpolated color values instead of a fixed background value or even skipping these additional points. In summary, the main objectives of this work are:

- **Propose a projection-based point cloud quality metric:** The proposed metric can handle both geometry and texture coding degradations by leveraging the work of the past decades on perceptual 2D quality metrics, such as MS-SSIM. This means that the proposed solution does not require any especially designed 2D quality metric and can accommodate any 2D quality metrics that will be proposed in the future.
- **Performance evaluation of available point cloud quality metrics:** The proposed metric is evaluated with a dataset that includes different types of coding degradations, that are created by different and relevant configurations of the recent standardized MPEG point cloud codecs. Moreover, its performance is evaluated in comparison to a large amount of state-of-the-art point cloud quality metrics, using different design approaches, but nevertheless achieves the top performance.

The last objective is very relevant towards achieving the objectives of this Thesis, since it will end with a performance evaluation with all the metrics proposed in the previous Chapters along with the other relevant state-of-the-art point cloud quality metrics, under relevant and meaningful test conditions.

The rest of this Chapter is organized as follows. Section 9.2 explains the proposed point cloud quality metric. Experimental results are presented and analyzed in Section 9.3.2 and finally, some final remarks are presented in Section 9.4.

## 9.2 Proposed Projection Based Point Cloud Quality Metric

In this section, the architecture of the proposed quality metric is presented and then the most relevant techniques are explained with more detail.

### 9.2.1 Architecture and Walkthrough

Figure 9.1 shows the proposed point cloud projection-based quality assessment metric architecture. The degraded and reference point clouds are processed in two parallel branches, resulting in two different intermediate scores which are fused at the end. To avoid any misalignment errors, before applying the 3D to 2D projection, both reference and decoded point clouds are processed to have exactly the same geometry. In the top branch, the geometry of the reference point cloud is used while on the bottom branch the geometry of the degraded point cloud is used, only the color attributes assigned to the points will be different, reflecting the color with and without coding degradations. Thus, in the top branch, the reference point cloud geometry is recolored with the color of degraded point cloud and the point cloud obtained is compared with reference point cloud (naturally, including color). In the other branch, the degraded point cloud geometry is recolored with the color of the reference point cloud and the resulting point cloud is compared with degraded point cloud (naturally, including color).

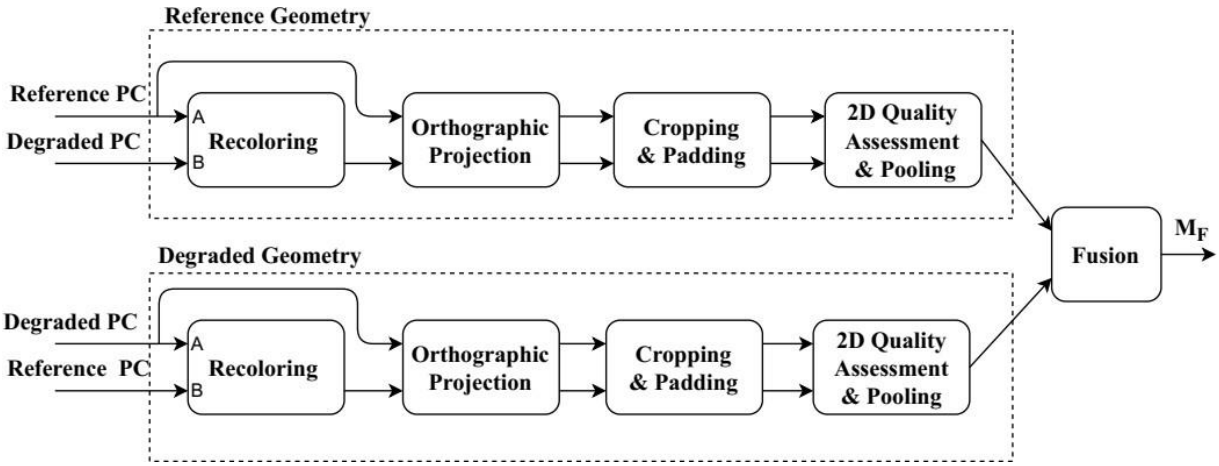



**Figure 9.1:** Block diagram of the proposed projection-based point cloud quality assessment metric.

Before applying this metric, point clouds should be voxelized if they are in floating point precision to some fixed precision. This step is important to perform the 3D (voxels) to 2D (pixels) projection. Nowadays, most of the point cloud data available is already in fixed precision (and was previously voxelized) and thus this step is not shown in the projection-based point cloud quality assessment architecture. Also, both V-PCC and G-PCC codecs code fixed precision point clouds or perform voxelization as a pre-processing step before coding. The main modules of the proposed point cloud quality metric pipeline are explained in the following.

1. **Recoloring:** Due to lossy coding of geometry data, the positions of the points in the reference and degraded point clouds are not the same. This is not an issue for point-based and feature-based metrics which find their correspondence by searching the nearest neighbor of every decoded point in the reference point cloud (using 3D coordinates) and vice versa. However, when a projection is made the resulting reference and degraded 2D images have regions that are not aligned, even when the degraded point cloud has only slightly degradations in geometry. 2D quality metrics are not very robust to misalignments (or displacements) and often perform poorly since pixel-wise comparisons are typically involved. To align both reference and degraded 2D images after projection, a recoloring step is applied where the decoded (or reference) point cloud color attributes are mapped to the reference (or decoded) geometry. In this way, the degradation of the color attributes is measured directly for two geometry levels (reference and decoded), thus considering both color and geometry. This solution is also rather suitable since color (or texture) degradations have a higher perceptual impact than geometry degradations. More details about this module are presented in section 9.2.2.
2. **Orthographic Projection:** The reference, degraded and the two recolored point clouds obtained from the previous step are orthographically projected on six faces of a cube to obtain six projected images for each, this means six different non-overlapping images for each point cloud. Although other projections could be used, the simplicity of this orthographic projection is enough to measure coding distortions, especially considering that the pairs of point clouds to compare have the same geometry. In this process, six occupancy maps are also obtained; an occupancy map is a binary map to indicate if a pixel corresponds to a point or not of the point cloud. The size of the projected texture and occupancy maps only depend on the precision $p$ of the point cloud $2^p \times 2^p$. More details about this module are presented in Section 9.2.29.2.3.
3. **Cropping:** After projection, depending on the size and position of the point cloud, the background may have a large area comparing to the point cloud object area. These empty areas can act as a distractor for the 2D quality assessment since the same color value is assigned to all background pixels in all projected point cloud images. More details about this module are presented in section 9.2.2.
4. **Padding:** A decoded point cloud may have a lower or higher number of points depending on the point cloud compression solution, e.g., when a point cloud is coded with octree pruning, a lower number of decoded points is obtained and when coded with V-PCC a higher number of points are typically obtained. When the number of points is different between the reference and degraded point clouds, the same pixel position may be occupied in one projected image while non-occupied in the other. In this situation, these pixel positions may significantly increase the error (i.e., account more distortion) when perceptually, this error is not visible. The padding process outputs a seamless image, where the background positions are filled with interpolated values, and thus closer to the image being shown after point cloud rendering. In this process, holes inside the point cloud are padded as well as any remaining empty space around objects. More details about this module are presented in section 9.2.2.

5. **2D Quality Assessment and Pooling:** A 2D image quality metric is computed between six reference images and the corresponding degraded images of the same view. The output of this process is six objective scores, one for each pair of projected images, which must be pooled together. The performance of the quality metric for typical pooling functions (max, min, weighted average, etc.) was evaluated and it was found that the performance is similar for different pooling functions. Thus, the resulting objective scores are averaged together (average pooling) to obtain a single score for each branch of the proposed projection-based quality metric.
6. **Fusion:** All modules previously described are applied to the two branches of the projection quality metric to obtain two intermediate scores. These two scores represent the distortion associated to the texture measured by a 2D quality metric, for two different geometry levels, i.e., for the reference geometry and for the degraded geometry and must be fused together to obtain the final metric. Although different fusion strategies are possible, even applying machine learning techniques, it was found that a simple linear regression was enough to obtain a high performance, without the risk of overfitting (in point clouds there is not many training data). More details about this process are presented in section 9.2.5.

### 9.2.2 Recoloring

The main challenge with projection-based metrics, is that geometry distortions may cause displacement errors between the projected reference and degraded 2D images after projection. However, 2D quality metrics do not handle well local displacement errors (or geometry errors) since pixel-wise comparisons are often made. When the same pixel location in the projected image is compared, the measured error may not reflect the perceived distortion since the color of different 3D locations in the reference and degraded point clouds are used in the error computation. Figure 9.2 shows frontal projection of Egyptian Mask before coding and after decoding with lowest geometry rate in G-PCC recolored with original color in (a) and (b) and their residual in c. Residual image is enhanced to better show the differences. Although color in both point clouds is equal and point clouds are visually similar, residual image is result of local displacement.

a) b) c)

**Figure 9.2:** *Egyptian Mask* projected from front view; a) Reference, b) Decoded (lowest geometry rate, G-PCC), c) Residual after image enhancement. Both reference and decoded point clouds have reference color.

Thus, the idea is to only use the reference or the degraded point cloud geometry and perform recoloring to assign the decoded color attributes to the reference geometry (top branch of the architecture) and the reference color attributes to the decoded geometry (bottom branch of the architecture). Using this approach, for each branch, the projected texture maps are aligned. To recolor point cloud *A*, with color of point cloud *B*, each point in point cloud *A* will have a color assigned using the color of one or more corresponding points in point cloud *B*. In proposed algorithm, the new color for each point of the point cloud *A* after recoloring is computed as follows:

1. For each point in point cloud *A*, the nearest neighbor in point cloud *B* is found ($NN_A$)
2. For each point in point cloud *B*, the nearest neighbor in point cloud *A* is found ($NN_B$)
3. Then, for each point $a$ in point cloud *A*:
    a. If point $a$ is listed in the nearest neighbors of point cloud *B* ($a \in NN_B$), the color is the average color of the points in point cloud *B* that have point $a$ as their nearest neighbor.

$$C_a = \sqrt{\frac{\sum_{b \in B, NN_B(b)=a} C_b^2}{\sum_{b \in B, NN_B(b)=a} 1}} \tag{9.1}$$

    where $C_a$ and $C_b$ are colors at points $a$ and $b$.

4. Otherwise, the color is the color of its nearest neighbor listed in $NN_A$.

Figure 9.3(a) shows the *Amphoriskos* reference point cloud (reference color and geometry) on the left and the recolored point cloud (decoded color with reference geometry) on the right. Figure 9.3(b) shows the *Amphoriskos* degraded point cloud (decoded color and geometry) on the left and the recolored point cloud (with reference color and decoded geometry) on the right. For both cases, the G-PCC codec with octree geometry coding mode and lifting color coding mode, in the lowest rate was used to obtain the degraded point cloud. A point-based rendering solution was used where point size in all point clouds is increased to fill empty spaces among points in decoded point cloud. Points are also represented with a cube.

(a) (b)

**Figure 9.3:** *Amphoriskos* decoded with G-PCC in octree mode and lifting color mode in the lowest rate; a) point clouds obtained for reference geometry with the reference (left) and decoded color, after recoloring (right), b) point clouds obtained for decoded geometry with the decoded color (left) and reference color, after recoloring (right).

### 9.2.3 Orthographic Projection

This is the core module of the proposed projection-based metric where 3D point clouds are mapped on several 2D planes from different perspectives. The orthographic projection is used for the projection-based metric because it is a parallel projection that preserves shapes and sizes. The orthographic projection, projects each point cloud point to a pixel in 2D image.

The proposed orthographic projection consists of two main steps:

- **Projection:** in this step, the planar (or 2D) images are generated using the orthographic projection to the different sides (planes) of the bounding box surrounding the object. For some plane, a point of the point cloud is projected into the plane as long as the point is not occluded by another point closer to the plane.
- **Filtering:** points that are projected to a plane but do not belong to the surface closer to the plane used for projection are removed. These points are projected when there is some empty space between points in the surface closer to the plane and therefore, points from other side of the object are projected from this plane. Since these points are not visible when rendering is performed, a filtering technique is used to remove these points.

Figure 9.4 shows a *Amphoriskos* before removing point from the back on top and after removing them on bottom, from two different views. The point cloud is recolored with 4 colors to better show the front and back part of the vase object from each view.

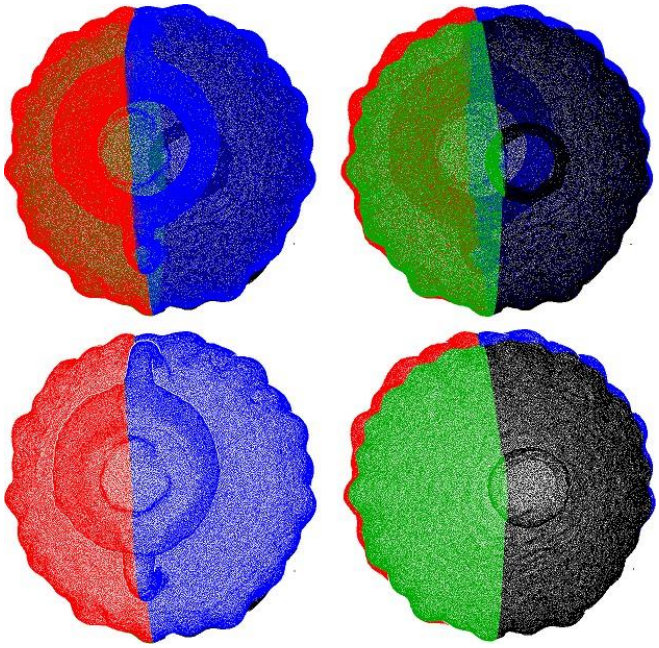


**Figure 9.4:** Projected point clouds from two different views: on top before filtering back parts and on bottom after filtering.

The orthographic projection algorithm projects every point that is visible from six planes around the object, if they are visible from that plane. If any point is closer to plane than an already projected point, it will overwrite that point. The proposed orthographic projection on six views proceeds as follows:

1. Six images, each for one of the projection planes are initialized with a uniform background color (white). Background color can be any color since further in the process of the proposed metric it will be filled with interpolated values. Three planes pass from origin: $\{xy, xz, yz\}$, and three pass from $2^p$: $\{\acute{x}\acute{y}, \acute{x}\acute{z}, \acute{y}\acute{z}\}$. Where $p$ is the precision of the point cloud

$$I_{xy} = I_{xz} = I_{yz} = I_{\acute{x}\acute{y}} = I_{\acute{x}\acute{z}} = I_{\acute{y}\acute{z}} = \begin{bmatrix} 255 & \cdots & 255 \\ \vdots & \ddots & \vdots \\ 255 & \cdots & 255 \end{bmatrix}_{2^p \times 2^p \times 3} \tag{9.2}$$

2. Six binary images are also initialized with zero as occupancy map (*ocm*) for each projected image.

$$ocm_{xy} = ocm_{xz} = ocm_{yz} = ocm_{\acute{x}\acute{y}} = ocm_{\acute{x}\acute{z}} = ocm_{\acute{y}\acute{z}} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{2^p \times 2^p} \tag{9.3}$$

3. To keep track of the occluded points, for each of the coordinates two matrix with the same size as images are initialized, one with minimum projected depth (*minDepth*), which is 0 and the

other with maximum depth (*maxDepth*) which is $2^p$. The minimum depth keeps the record of the depth of the last projected point on ($\acute{x}\acute{y}$, $\acute{x}\acute{z}$, $\acute{y}\acute{z}$) images and the maximum depth keeps the record of the depth of the last projected point on ($xy$, $xz$, $yz$) images.

$$minDepth_{xy} = minDepth_{xz} = minDepth_{yz} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{2^p \times 2^p} \tag{9.4}$$

$$maxDepth_{\acute{x}\acute{y}} = maxDepth_{\acute{x}\acute{z}} = maxDepth_{\acute{y}\acute{z}} = \begin{bmatrix} 2^p & \cdots & 2^p \\ \vdots & \ddots & \vdots \\ 2^p & \cdots & 2^p \end{bmatrix}_{2^p \times 2^p} \tag{9.5}$$

4. For each point $P_i = (P_x, P_y, P_z)$ of the point cloud with color $C_p$:

5. For each plane $K \in \{(xy, \acute{x}\acute{y}), (xz, \acute{x}\acute{z}), (yz, \acute{y}\acute{z})\}$ :

   i. Find the depth of point P, $P_d$ where d is the perpendicular axis to both planes in $K$.

   ii. If depth of point $P$, $P_d$ is less than or equal to the maximum depth at position $(P_u, P_v)$ of the plane $K$, then the point is projected on position $(P_u, P_v)$ of the image $I_{K1}$, corresponding pixel in occupancy map is also set to one and maximum depth is updated to the depth of $P_d$.

$$\text{if } P_d \leq maxDepth_{K1}(P_u, P_v) \text{ then} \tag{9.6}$$

$$I_{K1}(P_u, P_v) = C_p \text{ and } maxDepth_{K1}(P_u, P_v) = P_d \text{ and } ocm_{K1}(P_u, P_v) = 1 \tag{9.7}$$

   iii. If depth of point $P$, $P_d$ is larger than or equal to the minimum depth at position $(P_u, P_v)$ of the plane $K$, then the point is projected on position $(P_u, P_v)$ of the image $I_{K2}$, corresponding pixel in occupancy map is also set to one and minimum depth is updated to the depth of $P_d$.

$$\text{if } P_d \geq minDepth_{K2}(P_u, P_v) \text{ then} \tag{9.8}$$

$$I_{K2}(P_u, P_v) = C_p \text{ and } minDepth_{K2}(P_u, P_v) = P_d \text{ and } ocm_{K2}(P_u, P_v) = 1 \tag{9.9}$$

At this stage, every point that is not occluded will be projected. In this case, a point that belongs to the surface farther to the projection plane is always occluded by at least one point closer to the surface. These points are filtered out by comparing their depth to the depth of their neighboring pixels in a window $w$. The algorithm for filtering the points from back part of the point cloud is as follows:

1. For each occupied pixel $(a, b)$ in the projected image on planes $K \in \{xy, xz, yz\}$, compare the *minDepth* with average of *minDepth* of the neighbors in a window of size $w \times w$ and if it was smaller more than a threshold, that point is removed by resetting occupancy map and changing the value to background value.

$$\text{if } minDepth_K(a,b) \leq \frac{\sum_{(ij)\in w} minDepth_K(ij)}{\sum_{(ij)\in w} ocm_K(ij)} - \tau \text{ then} \tag{9.10}$$

$$l_K(a,b) = 255 \text{ and } ocm_K(a,b) = 0 \tag{9.11}$$

- For each occupied pixel $(a, b)$ in the projected image on planes $K \in \{xy, xz, yz\}$ compare the *maxDepth* with average of *maxDepth* of the neighbors in a window of size $w$ and if it was larger more than a threshold, that point is removed by resetting occupancy map and changing the value to background value.

$$\text{if } maxDepth_K(a,b) \geq \frac{\sum_{(ij)\in w} maxDepth_K(ij)}{\sum_{(ij)\in w} ocm_K(ij)} + \tau \text{ then} \tag{9.12}$$

$$l_K(a,b) = 255 \text{ and } ocm_K(a,b) = 0 \tag{9.13}$$

This is a proximity filtering algorithm that depends on the value of a threshold. This threshold Thus, it may also filter out a few of the points that belong to the front surface, but those points are later compensated in padding process described in the next section.

### 9.2.4 Cropping and Padding

After projecting a voxelized point cloud with precision of $p$, six images of size $2^p \times 2^p$ are obtained. Based on the size and shape of the object, these images contain a significant amount of background data, which comes from the empty spaces around the object. There might be also some amount of background pixels among the object region based on the density of the point cloud. In this work, the first group is removed by cropping out the Region of Interest (ROI), which is a box surrounding the object in that specific view and then the rest of empty spaces are filled by interpolated pixels obtained by the image inpainting technique.

To extract the ROI, the occupancy map for each projected image is used. More precisely, the position of the first and last occupied pixel in each occupancy map is used to define the bounding box for each object in the projected map.

To fill the holes and background pixels inside the projected object as well as the remaining backgrounds around the object, an image inpainting technique called Navier-Strokes [129] is used. Occupancy maps that are acquired during projection are used as a mask to guaranty that the occupied pixels will not change. Figure 9.5 shows an example of removing background process. *Longdress* is projected with a gray background to better show the image boundaries in Figure 9.5( a) shows the bounding box in green. This image is 1024×1024 since *Longdress* is 10-bit. In Figure 9.5b, the cropped ROI and in Figure 9.5 c, the padded image is shown.

(a) (b) (c)

**Figure 9.5:** *Longdress* decoded with G-PCC in octree mode and RAHT color coding mode at a medium rate, projected from frontal view: a) full size image including all voxels in 10-bit precision, with ROI in green; b) cropped ROI; c) padded image.

### 9.2.5 Fusion

As explained in Section 9.2.1, the proposed projection-based quality metric computes two different intermediate scores in two parallel branches. To combine two intermediate scores, the following linear model is used:

$$M_F = \alpha M_{Reference} + \beta M_{Degraded} \tag{9.14}$$

The $\alpha$ and $\beta$ parameters were estimated by an ordinary least squares linear regression procedure that minimizes the residual sum of squares between the objective scores predicted by the linear approximation and the mean opinion scores (MOS) available in some dataset. Although other complex models are possible to apply, they require more parameters and the risk of overfitting for the dataset to be used, this is rather critical since there are not many datasets with subjective scores and representative geometry and color degradations, especially compared to standard image and video subjective datasets.

Even for this linear model, to check that overfitting does not occur, data is randomly split to 75% training data and 25% test data for 100 times. For each iteration, $\alpha$ and $\beta$ values are estimated from the training data split and used to compute the final metric score on the test data split. The corresponding performance measured through the Pearson linear correlation coefficient (PLCC) is calculated for each iteration for the training data and is shown in Figure 9.6 ($M_F$). The average performance in all iterations (i.e., for different splits) is computed as the average PLCC over all iterations and also shown in Figure 9.6 (Average $M_F$). The PLCC performance calculated using all data for training and testing is also shown in Figure 9.6 (Final $M_F$). The final $M_F$ is very close

to the average $M_F$, which shows that the improvement in metric after fusion is not due to overfitting.

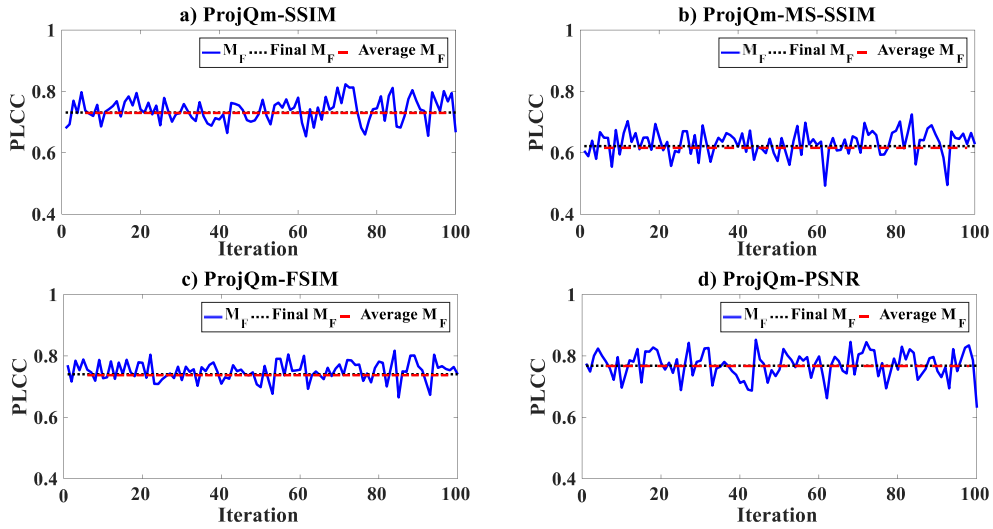



**Figure 9.6:** Proposed projection metric performance measured with the PLCC for every test set, average PLCC over all iterations and the PLCC using all data for training and testing. The PLCC is computed for different 2D quality metrics: a) SSIM, b) MS-SSIM, c) FSIM, d) PSNR.

## 9.3 Performance Evaluation

In this section, the proposed quality metrics performance is assessed, notably in comparison with state-of-the-art metrics, using point clouds coded with different type of point cloud codecs. Previous chapters were compared mostly with popular state-of-the-art quality metrics MPEG D1 and D2. Recently, many high-performance point cloud quality metrics are proposed in the literature. In this chapter, the projection quality metric is compared with all previous proposed metrics in this Thesis and also other relevant state-of-the-art point cloud quality metrics.

### 9.3.1 Subjective Evaluation Dataset

To perform this evaluation, the MPEG Point Cloud Compression Dataset (M-PCCD), publicly available in [130], have been used. This recent dataset includes 232 stimuli where color and geometry are both encoded, which is rather suitable for the experiments of this Chapter. The dataset includes both the MOS values as well as the reference and decoded point clouds.

The test material of this dataset corresponds to nine point clouds, including four objects and five human figures. *Longdress*, *Loot*, *Redandblack*, *Soldier*, *The20smaria* and *Head* from MPEG repository [104], *Romanoillamp* and *Biplane* from JPEG repository [131] and *Amphoriskos* from *Sketchfab* [132] as shown in Figure 9.7. *Redandback* is also used for training in this dataset. Characteristics of these point clouds are also listed in Table 9.1.

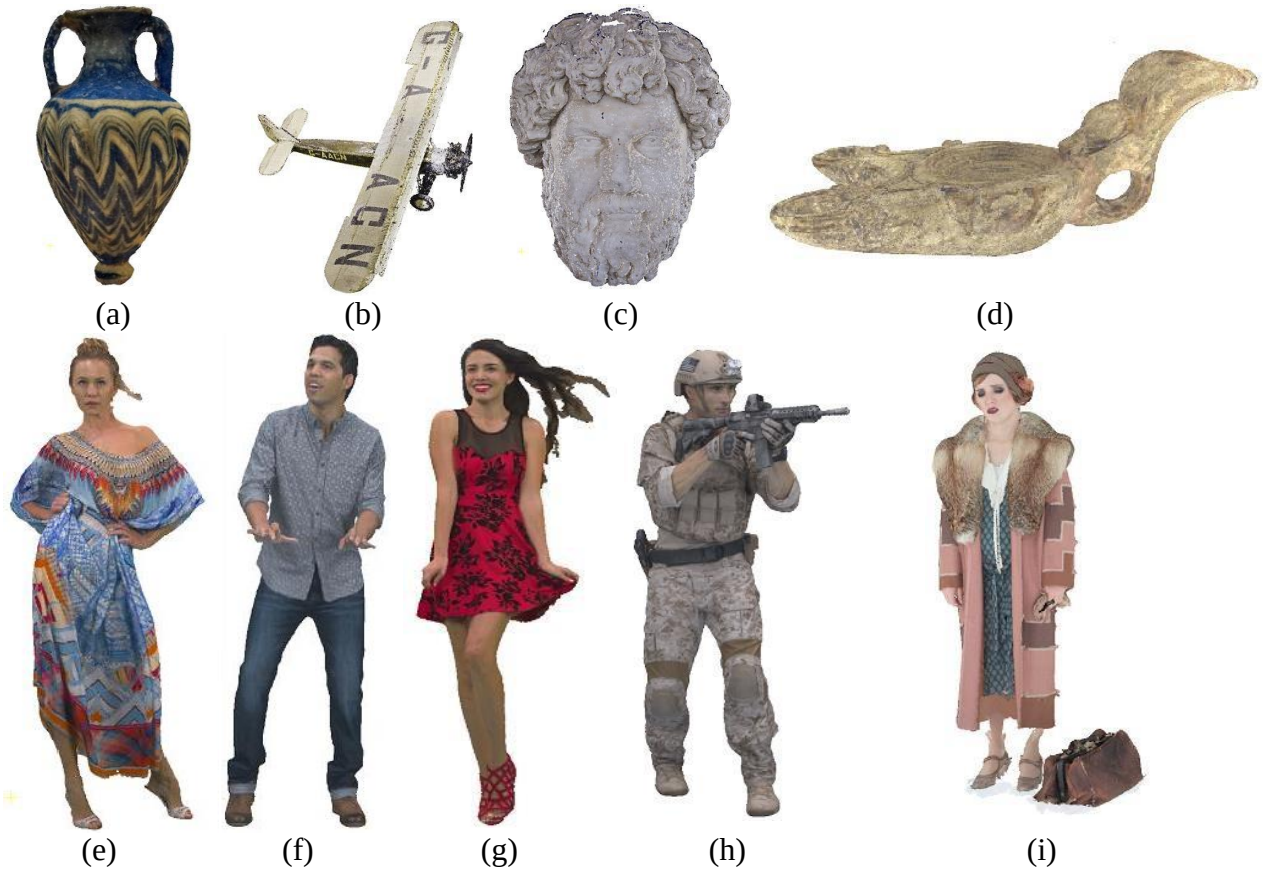



**Figure 9.7:** Test materials in M-PCCD. From top lef: a) *Amphoriskos, b) Biplane, c) Head, d) Romanoillamp, e) Longdress, f) Loot, g) Redandblack, h) Soldier, i) The20smaria.*

**Table 9.1:** Test point clouds in M-PCCD and associated characteristic.

| Name | Type | Repository | Precision | No. Points |
|---|---|---|---|---|
| *Amphoriskos* | Objects | *Sketchfab* | 10-bit | 814.474 |
| *Romanoillamp* | Objects | JPEG repository | 10-bit | 636.127 |
| *Biplane* | Objects | JPEG repository | 10-bit | 1.181.016 |
| *Head* | Objects | MPEG repository | 9-bit | 938.112 |
| *Longdress* | People | MPEG repository | 10-bit | 857.966 |
| *Loot* | People | MPEG repository | 10-bit | 805.285 |
| *Redandblack* | People | MPEG repository | 10-bit | 757.691 |
| *Soldier* | People | MPEG repository | 10-bit | 1.089.091 |
| *The20smaria* | People | MPEG repository | 10-bit | 1.553.937 |

The point clouds have been coded in six rates with MPEG G-PCC with octree geometry encoding module and RAHT color coding module, six rates with MPEG G-PCC with octree geometry encoding module and Lifting color coding module, with MPEG G-PCC with TriSoup geometry encoding module and RAHT color coding module, six rates with MPEG G-PCC with TriSoup geometry encoding module and Lifting color coding module and finally in five rates with MPEG V-PCC codec. The rates were selected based on the suggestion on MPEG CTC [104].

**TABLE 9.2:** Quantization parameters for geometry and color for both codec, and octree level for G-PCC octree and trisoup

| | | G-PCC | | | | V-PCC |
|---|---|---|---|---|---|---|
| | | **TriSoup** | **TriSoup** | **Octree** | **Octree** | |
| **Rate** | | **RAHT** | **Lifting** | **RAHT** | **Lifting** | |
| **R6** | Color QP | 8 | 8 | 8 | 8 | - |
| | Position QP | 15/16 | 15/16 | 15/16 | 15/16 | - |
| | level | Depth | Depth | Depth | Depth | - |
| **R5** | Color QP | 16 | 16 | 16 | 16 | 22 |
| | Position QP | 7/8 | 7/8 | 7/8 | 7/8 | 16 |
| | level | Depth-2 | Depth-2 | Depth-2 | Depth-2 | - |
| **R4** | Color QP | 32 | 32 | 32 | 32 | 27 |
| | Position QP | 3/4 | 3/4 | 3/4 | 3/4 | 20 |
| | level | Depth-3 | Depth-3 | Depth-3 | Depth-3 | - |
| **R3** | Color QP | 64 | 64 | 64 | 64 | 32 |
| | Position QP | 1/2 | 1/2 | 1/2 | 1/2 | 24 |
| | level | Depth-3 | Depth-3 | Depth-3 | Depth-3 | - |
| **R2** | Color QP | 128 | 128 | 128 | 128 | 37 |
| | Position QP | 1/4 | 1/4 | 1/4 | 1/4 | 28 |
| | level | Depth-3 | Depth-3 | Depth-3 | Depth-3 | - |
| **R1** | Color QP | 256 | 256 | 256 | 256 | 42 |
| | Position QP | 1/8 | 1/8 | 1/8 | 1/8 | 32 |
| | level | Depth-4 | Depth-4 | Depth-4 | Depth-4 | - |

A subjective study with DSIS methodology in performed in two separate labs each with 20 subjects. Point clouds are rendered with *RPoint*, and point clouds were shown side-by-side in an interactive evaluation protocol allowing subjects to modify their viewpoint.

The outlier detection algorithm described in the ITU-R Recommendation BT.500-13 [31] was issued separately for each laboratory, in order to exclude subjects whose ratings deviated

drastically from the rest of the scores. As a result, no outliers were identified, thus, leading to 20 ratings per stimulus at each lab. Then MOS is computer by averaging scores between all subjective for each content.

### 9.3.2 Experimental Results and Analysis

In this section, objective-subjective correlation performance of the proposed metric is evaluated and compared with some benchmarks, as defined in previous Chapters and in the literature. However, before evaluating the point cloud quality metrics performance, nonlinear fitting is applied on the objective scores, namely three different regressors: a monotonic cubic function (9.15), a logistic function (9.16) and a quadratic function (9.17).

$$MOS_p = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 \tag{9.15}$$

$$MOS_p = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-\left(\frac{Q_i - \beta_3}{\beta_4}\right)}} \tag{9.16}$$

$$MOS_p = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 \tag{9.17}$$

where $x$ are objective metric values and $\beta_1, \ldots, \beta_4$ are the regression model parameters. This approach allows to fit the objective metric values to the perceptual (MOS) scale and thus, obtain the predicted MOS scores. To assess the performance of any point cloud quality metric, the PLCC, SROCC and RMSE are used. When the values of PLCC and SROCC are close to one, the predicted objective scores are highly correlated and have a monotonic relationship with the ground-truth MOS values. Moreover, SROCC does not depend on the selected fitting function since it is a measure of monotonicity.

**Benchmarks**

List of the metrics that we have compared proposed metric with them are as follows:

- **D1 and D1-PSNR** [104]**:** Po2Po MSE error and PSNR associated to this error.
- **D2 and D2-PSNR** [104]**:** Po2Pl MSE error and PSNR associated to this error.
- **Hausdorff distance and PSNR** [133]**:** Po2Po and Po2Pl Hausdorff distance and PSNR associated to this error.
- **Y-MSE and Y-PSNR** [104]**:** mean square error between luminance of points and their nearest neighbor and PSNR associated to this error.
- **Pl2Pl MSE** [109]**:** mean squared of angular (cosine) similarity between underlying surfaces at each point and its nearest neighbor.
- **PCQM** [134]**:** three geometry features related to curvature are combined with five color features related to lightness, chroma and hue. PCQM corresponds to the weighted average of the differences for geometry and color features between reference and decoded point clouds.

- **Point SSIM** [135]**:** The features are extracted in a local neighborhood around each point in the reference and degraded point clouds considering four independent 'attributes', notably geometry, color (luminance), normal and curvature information. Results for the best variant is used which is considering only color features, neighborhood size (k) of 12, original voxel depth and using variance as statistical dispersion measure.
- $\boldsymbol{H}_{l2}^{Y}$ [136]**:** A point cloud quality metric based on the histogram and correlogram of the luminance component. Raw scores published by the Author are used in this Thesis.
- $\boldsymbol{d_{gc}}$ [136]**:** A linear combination of Po2Pl MSE with $H^{Y}_{l2}$. Raw scores published by the Author are used in this Thesis.
- **PCM$_{RR}$** [137]**:** A reduced reference quality metric that jointly evaluates color and geometry. A set of seven statistical features such as mean, standard deviation, etc. are extracted from the reference and degraded point clouds in the geometry, texture, and normal vector domain, in a total of 21 features. The quality score is computed as the weighted average of the differences for all these features between the reference and degraded point clouds, with the weights obtained through a linear optimization algorithm. Raw scores of the best variant published by the Author are used in this Thesis.
- **Geotex** [138]**:** A metric based on Local Binary Pattern (LBP) descriptors adapted to point clouds and applied to the luminance. Histograms of the extracted feature maps are obtained for both the reference and the degraded point clouds to be compared using a distance metric such as f-divergence. Scores published on the article is used in this Thesis.
- **Projection-based metric** [126]**:** point clouds are projected on six faces of a cube. Six images are concatenated without any pre-processing and PSNR of luminance and RGB-PSNR are used as quality metric.
- **Projection-based metric** [121]**:** The same rendered point clouds from 42 views that are shown to the subjects in subjective evaluation are evaluated by different 2D metrics. The final value of metric is pooled using l1-norm.

And also, all proposed metrics in this Thesis as follows:

- **GH-PSNR:** The generalized Hausdorff distance-based PSNR proposed in Chapter 6, considering 98% of the distances, using maximum pooling function ($PSNR_{98,max}$). Both Po2Po and Po2Pl distances are used
- **RA-PSNR:** Resolution-adaptive PSNR proposed in Chapter 7. The rendering resolution of 10 nearest neighbors is used for the experiments of this Chapter ($APD_{10}$). This metric considers intrinsic resolution (considering the rendering impact) of points as well as precision.

- **P2D MMD-PSNR:** point-to-distribution metric proposed in Chapter 8, using only the PSNR of Mean Mahalanobis Distances (MMD). This metric finds the distance between a point a distribution of a point.

In this section, the performance of the proposed metric is compared with the best performing projection-based metrics available in the literature [121], see Section 9.3.2.1. The overall subjective-objective correlation performance of the projection-based quality metric is compared with other state-of-the-art point cloud quality metrics, in section 9.3.2.2. An ablation study is performed to show the module-by-module performance of the projection-based quality metric, in section 9.3.2.5.

### *9.3.2.1 Performance against State-of-the-Art Projection-based Metrics*

In this section, performance of the proposed metric is compared with the projection-based metrics proposed in [121], which were first proposed in [127]. The performance assessment of the proposed projection-based quality metric is done using the same 2D quality metric as those used in the benchmarks.

**Table 9.3:** Objective-subjective correlation performance of the proposed metric compared with **[121]** and **[70]** for same 2D quality metric.

| **METRIC** | | ***SROCC*** | **Cubic** | | **Logistic** | | **Quadratic** | |
|---|---|---|---|---|---|---|---|---|
| | | | ***PLCC*** | ***RMSE*** | ***PLCC*** | ***RMSE*** | ***PLCC*** | ***RMSE*** |
| VIFP | *Proposed* | **85.5** | **83.0** | **0.760** | **83.0** | **0.760** | **83.0** | **0.850** |
| | *Alexiou et al.* [121] | 74.2 | 71.6 | 0.949 | 71.5 | 0.951 | 71.4 | 0.953 |
| | *Gain* | 11.3 | 11.4 | 0.189 | 11.5 | 0.191 | 11.6 | 0.103 |
| SSIM | *Proposed* | **81.3** | **80.6** | **0.800** | **80.9** | **0.800** | **79.6** | **0.810** |
| | *Alexiou et al.* [121] | 63.3 | 63.6 | 1.049 | 62.6 | 1.061 | 62.1 | 1.066 |
| | *Gain* | 18.0 | 17.0 | 0.249 | 18.3 | 0.261 | 17.5 | 0.256 |
| MS-SSIM | *Proposed* | **82.8** | **77.0** | **0.870** | **79.5** | **0.830** | **73.4** | **0.820** |
| | *Alexiou et al.* [121] | 75.2 | 70.1 | 0.970 | 70.9 | 0.959 | 68.8 | 0.987 |
| | *Gain* | 7.6 | 6.9 | 0.100 | 8.6 | 0.129 | 4.6 | 0.167 |
| Y-PSNR | *Proposed* | **79.1** | **77.1** | **0.870** | **77.1** | **0.870** | **77.1** | **0.790** |
| | *Alexiou et al.* [121] | 62.8 | 62.6 | 1.061 | 66.7 | 1.013 | 59.7 | 1.091 |
| | *Gain* | 16.3 | 14.5 | 0.191 | 10.4 | 0.143 | 17.4 | 0.301 |
| | *de Queiroz et al* [126]. | 33.1 | 43.5 | 1.225 | 43.0 | 1.228 | 38.8 | 1.254 |
| | *Gain* | 46.0 | 33.6 | 0.355 | 34.1 | 0.358 | 38.3 | 0.464 |

Table 9.3 clearly shows that proposed projection-based quality metric outperforms the projection-based metrics benchmarks significantly, for the same 2D quality metric. The minimum and maximum gains of the proposed projected quality metric are 18.3 for PLCC, 18.0 for SROCC and 0.3 for RMSE, comparing to *Alexiou et al.* [121]:. These gains are rather high and consistent among quality metric performance measures (PLCC, SROCC and RMSE), especially considering that [37] uses 42 views, which is much higher (and also more complex) than the six views considered in the proposed metric. Other interesting conclusions is that in this limited study, the VIFP is the best 2D quality metric and there is no significant difference when different fitting functions are used, for VIFP and SSIM, while the differences for MS-SSIM and PSNR-Y are higher.

#### *9.3.2.2 Performance for 2D Quality Assessment Metrics*

As stated in Section 9.2, the proposed projection-based point cloud quality metric is flexible enough to accommodate any 2D quality metric. In this Section, the performance of the proposed projection-based quality metric is evaluated for different 2D quality metrics. This will allow to identify which 2D metric leads to the highest overall performance. In this case, the same 2D quality metric is used in both branches of the proposed metric architecture, especially because the fusion module works best when the quality range and scale is similar for both reference and degraded geometry branches. The 2D quality metric used can influence the performance of the projection-based metric significantly and thus a wide set of quality metrics were evaluated: 1) PSNR, 2) PSNR-HVS [139], 3) PSNR-HVS-M [140], 4) Structural Similarity Index Metric (SSIM) [141], 4) Multi-Scale Structural Similarity Index Metric (MS-SSIM) [120], 5) Visual Information Fidelity Measure (VIFP) [142], 6) Feature Similarity Index (FSIM) [143], 7) Visual Saliency Index (VSI) [144], 8) Learned Perceptual Image Patch Similarity (LPIPS) [145], 9) Deep Image Structure and Texture Similarity (DISTS) [146], 11) Haar Perceptual Similarity Index (HaarPSI) [147].

Table 9.4 shows the correlation performance of the proposed projected-based quality metric for a large set of 2D quality metrics, considering all possible coding degradations (all data). DISTS has the best performance among all the 2D quality metrics while LPIPS, FSIM, and VSI come in the following positions. Both DISTS and LPIPS are very recent 2D quality metrics that use powerful deep-learning features to perform quality assessment. More specifically, DISTS includes both the texture similarity and structure similarity components, which are weighted to achieve a better correlation with perceived quality and to be invariant to small changes in texture patches (homogenous regions with repeated elements). LPIPS computes distances between features computed for both reference and degraded input images at different layers of a neural network. In both DISTS and LPIPS metrics, pixel-wise comparisons are not performed, and a feature perceptual space is used instead. Typically, these metrics weight more general appearance changes rather than small changes in textures, where their elements may have different location, size, color and orientation. This actually fits rather well the projected images obtained by the proposed metric where small texture changes may occur due to the recoloring process. The projection-based quality

metric with the best four 2D quality metrics will be used in the remaining experiments of this chapter.

**Table 9.4:** Objective-subjective correlation performance of the proposed projection-based quality metric for different 2D quality metrics.

| 2D Metric | *SROCC* | Cubic | | Logistic | | Quadratic | |
|---|---|---|---|---|---|---|---|
| | | ***PLCC*** | ***RMSE*** | ***PLCC*** | ***RMSE*** | ***PLCC*** | ***RMSE*** |
| *DISTS* | **95.6** | **94.6** | **0.440** | **94.7** | **0.439** | **93.6** | **0.478** |
| *LPIPS* | 93.2 | 92.2 | 0.525 | 92.3 | 0.523 | 92.2 | 0.527 |
| *FSIM* | 90.1 | 87.6 | 0.655 | 88.2 | 0.640 | 85.4 | 0.708 |
| *HaarPSI* | 87.7 | 84.7 | 0.723 | 84.8 | 0.721 | 84.4 | 0.729 |
| *VSI* | 87.6 | 83.9 | 0.74 | 85.4 | 0.707 | 80.3 | 0.811 |
| *VIPF* | 85.5 | 83.0 | 0.758 | 83.0 | 0.758 | 83.0 | 0.758 |
| *MS-SSIM* | 82.8 | 77.0 | 0.869 | 79.5 | 0.825 | 73.4 | 0.923 |
| *SSIM* | 81.3 | 80.6 | 0.805 | 80.9 | 0.800 | 79.6 | 0.823 |
| *PSNR HVS M* | 81.3 | 78.3 | 0.846 | 78.7 | 0.840 | 78.0 | 0.852 |
| *PSNR HVS* | 80.5 | 78.3 | 0.845 | 78.4 | 0.845 | 78.2 | 0.848 |
| *Y-PSNR* | 79.1 | 77.1 | 0.866 | 77.1 | 0.866 | 77.1 | 0.866 |

#### *9.3.2.3 Performance against State-of-the-Art Point Cloud Quality Metrics*

In this section, the performance of the proposed projection-based quality metric is compared with, if not all, state-of-the-art metrics in the point cloud quality assessment literature. Table 9.5 shows the SROCC, PLCC and RMSE of several benchmark metrics in the literature, in the evaluation of the decoded point cloud quality, again considering all data. The following conclusions can be derived:

- **Overall performance:** the proposed projection-based metric using DISTS is the best performing metric for point cloud quality assessment with the highest PLCC, SROCC and lowest RMSE values. This is consistent for the several fitting functions used in this evaluation. This result also confirms that by projecting a point cloud into several 2D images (which are close to what a user sees) and by exploiting the power of 2D quality metrics a top result can be achieved. The proposed projection-based quality metric with LPIPS and the point SSIM quality metrics also have a very high performance.
- **Proposed vs point-based point cloud quality metrics:** The proposed projection-based metric significantly outperforms the point-based D1 and D2-PSNR and plane-to-plane quality metrics that are currently used by the MPEG and JPEG standardization groups. The gains are rather

significant, up to 11.8 points in SROCC and 19.5 points in PLCC. Moreover, comparing to the other point-based quality metrics a similar gain can be achieved.

- **Proposed vs feature-based point cloud quality metrics:** The feature-based metrics, however, show better performance than point-based metrics and thus the gains are not so high considered the proposed 2D quality metric. The overall performance of the best proposed projection-based metric is almost 4% in SROCC and 2% in PLCC higher than the best feature-based metric.
- **Proposed vs previously presented Thesis metrics:** Proposed projection-based metric shows the best performance comparing to all proposed metrics in this thesis. The next best proposed metrics are Po2Po RA-PSNR, P2D MMSD and GH-PSNR (98%).

**Table 9.5:** Objective-subjective correlation performance of the proposed metric, comparing with the point cloud quality assessment state-of-the-art. Cells with a dash (-) shows missing performance since some of these results were obtained from the corresponding paper.

| Type | Evaluation Domain | Metric Name | *SROCC* | *Cubic* | | *Logistic* | | *Quadratic* | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | *PLCC* | *RMSE* | *PLCC* | *RMSE* | *PLCC* | *RMSE* |
| *Point-based (Po2Po)* | *Geometry* | *D1* [104] | 86.8 | 78.5 | 0.843 | 84.8 | 0.722 | 70.5 | 0.965 |
| | *Geometry* | *D1-PSNR* [104] | 79.7 | 77.6 | 0.858 | 77.7 | 0.857 | 73.1 | 0.929 |
| | *Geometry* | *Hausdorff* [133] | 37.0 | 19.3 | 1.335 | 1.4 | 1.360 | 14.2 | 1.347 |
| | *Geometry* | *Hausdorff PSNR* [133] | 36.6 | 62.4 | 1.064 | 66.1 | 1.021 | 52.5 | 1.158 |
| | *Color + Geometry* | *Y-MSE* [104] | 66.2 | 55.1 | 1.135 | 66.3 | 1.018 | 49.2 | 1.184 |
| | *Color + Geometry* | *Y-PSNR* [104] | 66.2 | 67.0 | 1.010 | 67.1 | 1.009 | 67.0 | 1.010 |
| *Point-based (Po2Pl)* | *Geometry* | *D2* [104] | 88.4 | 75.0 | 0.899 | 85.9 | 0.695 | 68.9 | 0.986 |
| | *Geometry* | *D2-PSNR* [104] | 83.8 | 80.3 | 0.810 | 80.5 | 0.808 | 75.1 | 0.899 |
| | *Geometry* | *Hausdorff* [133] | 50.5 | 35.6 | 1.271 | 67.2 | 1.007 | 26.3 | 1.312 |
| | *Geometry* | *Hausdorff PSNR* [133] | 49.3 | 59.6 | 1.092 | 56.3 | 1.124 | 56.0 | 1.127 |
| *Point-based (Pl2Pl)* | *Geometry* | *MSE* [109] | 47.7 | 58.5 | 1.104 | 62.4 | 1.063 | 58.2 | 1.106 |
| *Feature-based* | *Geometry* | *PCQM* [134] | 91.6 | 88.7 | 0.629 | 89.9 | 0.597 | - | - |
| | *Geometry* | $d_{gc}$ [136] | 92.0 | 80.6 | 0.804 | 90.4 | 0.585 | 72.2 | 0.941 |
| | *Color + Geometry* | $H^{Y}_{L2}$ [136] | 88.4 | 82.0 | 0.78 | 85.3 | 0.710 | 74.7 | 0.905 |
| | *Color + Geometry* | $PCM_{RR}$*(MCCV)* [137] | 90.7 | 89.4 | 0.594 | 90.2 | 0.573 | 77.6 | 0.858 |
| | *Geometry, Color* | *Point SSIM* [135] | 91.8 | 92.5 | 0.516 | 92.6 | 0.514 | 91.2 | 0.558 |
| | *Color + Geometry* | *Geotex* [138] | 87.9 | 66.7 | 1.042 | - | - | - | - |
| *Projection-based* | *Color + Geometry* | *Y-MS-SSIM* [121] | 75.2 | 70.1 | 0.97 | 70.9 | 0.959 | 68.8 | 0.987 |
| | *Color + Geometry* | *Y-VIFP* [121] | 74.2 | 71.6 | 0.949 | 71.5 | 0.951 | 71.4 | 0.953 |
| | *Color + Geometry* | *Y-PSNR* [126] | 33.1 | 43.5 | 1.225 | 43.0 | 1.228 | 38.8 | 1.254 |
| *Point-based (Po2Po)* | *Geometry* | *GH 98% PSNR (Ch. 6)* | 86.9 | 84.6 | 0.725 | 84.6 | 0.726 | 79.8 | 0.820 |
| | *Geometry* | *RA-PSNR* ($APD_{10}$) *(Ch. 7)* | 90.2 | 87.8 | 0.651 | 88.8 | 0.626 | 79.0 | 0.833 |
| *Point-based (Po2Pl)* | *Geometry* | *GH 98% PSNR (Ch. 6)* | 87.9 | 84.3 | 0.732 | 84.3 | 0.731 | 79.7 | 0.822 |
| | *Geometry* | *RA-PSNR* ($APD_{10}$) *(Ch. 7)* | 89.9 | 88.8 | 0.625 | 88.9 | 0.622 | 80.7 | 0.803 |
| *Point-based (P2D)* | *Geometry* | *MMD (Ch. 9)* | 88.7 | 86.0 | 0.695 | 86.9 | 0.672 | 80.5 | 0.807 |
| | *Geometry* | *MMD-PSNR (Ch. 9)* | 88.7 | 86.7 | 0.679 | 86.9 | 0.673 | 86.7 | 0.679 |
| *Projection-based* | *Color + Geometry* | *ProjQm-FSIM (Ch. 10)* | 90.1 | 87.6 | 0.66 | 88.2 | 0.64 | 85.4 | 0.89 |
| | *Color + Geometry* | *ProjQm-VSI (Ch. 10)* | 87.6 | 83.9 | 0.74 | 85.4 | 0.707 | 80.3 | 0.811 |
| | *Color + Geometry* | *ProjQm-LPIPS (Ch. 10)* | 93.2 | 92.2 | 0.525 | 92.3 | 0.523 | 92.2 | 0.527 |
| | *Color + Geometry* | *ProjQm-DISTS (Ch. 10)* | **95.6** | **94.6** | **0.44** | **94.7** | **0.439** | **93.6** | **0.478** |

#### *9.3.2.4 Coding Approach Impact on the Proposed Projection-based Quality Metric*

The performance of the projection-based quality metric can be significantly influenced by the type of coding artifacts, which are generated by some specific type of point cloud coding, e.g., an octree codec typically produces very distinct artifacts such as pixelization or blocking artifacts, which can be easily identified. Therefore, the dataset has been split into two identifiable sets

- MPEG G-PCC: the point clouds that are decoded with G-PCC in six rates for TriSoup geometry coding mode and RAHT and Predicting color coding modes and six rates for Octree geometry coding mode with the same two color coding modes. G-PCC decoded point clouds for each test material is 24 rates in total.
- MPEG V-PCC: the point clouds that are decoded with G-PCC in six rates.

From this set, the decoded point clouds of MPEG V-PCC codec have always been challenging to point cloud quality assessment metrics. Table 9.6 shows the objective-subjective correlation performance of the proposed projection-based quality metric compared with the state-of-the-art point cloud quality metrics for V-PCC decoded point clouds. The proposed ProjQM-DISTS is the best performing metric for V-PCC decoded data with 85.3% of SROCC and more than 86% of PLCC. The point-based metrics and projection-based metrics in the state-of-the-art are failing to reliable assess the V-PCC decoded quality. Proposed metric using DISTS outperforms the best projection-based metric in the literature (Y-VIFP) by 49.7% SROCC and 41.3% PLCC. These numbers are 16.5% and 26.1 for point-based metrics (D2). For feature-based metrics that are the best metrics in the literature the proposed metric's gains is 0.8% SROCC and 3.4% PLCC comparing with the best metric (PointSSIM). However, proposed projection-based metric using DISTS 2D metric and PointSSIM from feature-based metrics are showing significantly better results comparing to other metrics.

The objective-subjective correlation performance results for G-PCC codec are presented in Table 9.7. The proposed projection-based metric using DISTS is the best performing metric for G-PCC decoded data with 96% of SROCC and more than 95.8% of PLCC. Feature-based and point-based metrics (except for Pl2Pl) are showing acceptable performance assessing the G-PCC decoded data. However, proposed projection-based metric using DISTS outperforms the best point-based metric (D2) by 5.4% SROCC and 7.9% PLCC. The correlation gains against the best feature-based metric ($d_{gc}$) are 2.1% SROCC and 3.3% PLCC. State-of-the-art projection-based metric does not show an acceptable performance evaluating G-PCC decoded point clouds.

**Table 9.6:** Objective-subjective correlation performance of the proposed projection-based metric, comparing with the point cloud quality assessment state-of-the-art, for V-PCC. Cells with a dash (-) shows missing performance since some of these results were obtained from the corresponding paper.

| Type | Evaluation Domain | Metric Name | *SROCC* | *Cubic* | | *Logistic* | | *Quadratic* | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | ***PLCC*** | ***RMSE*** | ***PLCC*** | ***RMSE*** | ***PLCC*** | ***RMSE*** |
| *Point-based (Po2Po)* | *Geometry* | *D1* [104] | 42.0 | 42.0 | 0.950 | 46.3 | 0.928 | 41.6 | 0.952 |
| | *Geometry* | *D1-PSNR* [104] | 28.2 | 35.0 | 0.981 | 30.4 | 0.997 | 30.3 | 0.998 |
| | *Geometry* | *Hausdorff* [133] | -17.5 | 21.2 | 1.023 | 14.7 | 1.047 | 20.6 | 1.024 |
| | *Geometry* | *Hausdorff PSNR* [133] | -14.9 | 37 | 0.973 | 27.1 | 1.008 | 15.8 | 1.034 |
| | *Color + Geometry* | *Y-MSE* [104] | 33.3 | 49.5 | 0.909 | 37.9 | 0.969 | 19.0 | 1.028 |
| | *Color + Geometry* | *Y-PSNR* [104] | 33.3 | 39.1 | 0.964 | 37.6 | 0.97 | 34.9 | 0.981 |
| *Point-based (Po2Pl)* | *Geometry* | *D2* [104] | 68.8 | 71.3 | 0.734 | 73.5 | 0.71 | 71.0 | 0.737 |
| | *Geometry* | *D2-PSNR* [104] | 55.3 | 53.9 | 0.882 | 60.3 | 0.835 | 53.9 | 0.882 |
| | *Geometry* | *Hausdorff* [133] | 12.8 | 20.5 | 1.025 | 23.8 | 1.017 | 16.6 | 1.032 |
| | *Geometry* | *Hausdorff PSNR* [133] | 13.5 | 16.2 | 1.033 | 28.6 | 1.003 | 16.2 | 1.033 |
| *Point-based (Pl2Pl)* | *Geometry* | *MSE* [109] | 34.1 | 54.4 | 0.879 | 51.6 | 0.897 | 48.5 | 0.916 |
| *Feature-based* | *Geometry* | *PCQM* [134] | - | - | - | - | - | - | - |
| | *Geometry* | $d_{gc}$ [136] | 74.0 | 74.9 | 0.694 | 75.3 | 0.689 | 74.4 | 0.700 |
| | *Color + Geometry* | $H^{Y}_{L2}$ [136] | 68.3 | 65.9 | 0.787 | 65.7 | 0.789 | 65.3 | 0.793 |
| | *Color + Geometry* | $PCM_{RR}$*(MCCV)* [137] | 64.8 | 71.9 | 0.727 | 71.6 | 0.731 | 67.3 | 0.774 |
| | *Geometry, Color* | *Point SSIM* [135] | 84.5 | 83.0 | 0.584 | 83.0 | 0.584 | 82.8 | 0.587 |
| | *Color + Geometry* | *Geotex* [138] | - | - | - | - | - | - | - |
| *Projection-based* | *Color + Geometry* | *Y-MS-SSIM* [121] | 35.4 | 45.5 | 0.932 | 31.9 | 0.992 | 31.1 | 0.995 |
| | *Color + Geometry* | *Y-VIFP* [121] | 35.6 | 34.8 | 0.982 | 43.7 | 0.942 | 34.5 | 0.983 |
| | *Color + Geometry* | *Y-PSNR* [126] | -10.1 | 31.9 | 0.992 | 29.7 | 1.000 | 21.4 | 1.024 |
| *Point-based (Po2Po)* | *Geometry* | *GH 98% PSNR (Ch. 6)* | 57.3 | 59.3 | 0.843 | 57.8 | 0.854 | 53.4 | 0.885 |
| | *Geometry* | *RA-PSNR* ($APD_{10}$) *(Ch. 7)* | 67.3 | 70.1 | 0.747 | 68.9 | 0.759 | 63.3 | 0.810 |
| *Point-based (Po2Pl)* | *Geometry* | *GH 98% PSNR (Ch. 6)* | 71.2 | 75.4 | 0.687 | 75.1 | 0.691 | 75.0 | 0.693 |
| | *Geometry* | *RA-PSNR* ($APD_{10}$) *(Ch. 7)* | 76.9 | 80.3 | 0.625 | 79.9 | 0.629 | 77.0 | 0.668 |
| *Point-based (P2D)* | *Geometry* | *MMD (Ch. 9)* | 69.0 | 72.1 | 0.725 | 71.8 | 0.729 | 70.9 | 0.738 |
| | *Geometry* | *MMD-PSNR (Ch. 9)* | 69.0 | 72.4 | 0.722 | 71.9 | 0.728 | 70.1 | 0.747 |
| *Projection-based* | *Color + Geometry* | *ProjQm-FSIM (Ch. 10)* | 72.1 | 71.3 | 0.734 | 71.0 | 0.737 | 71.2 | 0.737 |
| | *Color + Geometry* | *ProjQm-VSI (Ch. 10)* | 63.9 | 64.2 | 0.802 | 64.5 | 0.800 | 64.2 | 0.802 |
| | *Color + Geometry* | *ProjQm-LPIPS (Ch. 10)* | 79.5 | 80.5 | 0.621 | 80.7 | 0.618 | 79.9 | 0.630 |
| | *Color + Geometry* | *ProjQm-DISTS (Ch. 10)* | **85.3** | **86.2** | **0.532** | **86.4** | **0.526** | **85.7** | **0.540** |

**Table 9.7:** Objective-subjective correlation performance of the proposed projection-based metric, comparing with the point cloud quality assessment state-of-the-art, for G-PCC. Cells with a dash (-) shows missing performance since some of these results were obtained from the corresponding paper.

| **Type** | **Evaluation Domain** | **Metric Name** | ***SROCC*** | ***Cubic*** | | ***Logistic*** | | ***Quadratic*** | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | ***PLCC*** | ***RMSE*** | ***PLCC*** | ***RMSE*** | ***PLCC*** | ***RMSE*** |
| *Point-based (Po2Po)* | *Geometry* | *D1* [104] | 90.0 | 81.5 | 0.814 | 88.6 | 0.651 | 73.0 | 0.960 |
| | *Geometry* | *D1-PSNR* [104] | 83.9 | 82.1 | 0.801 | 82.5 | 0.794 | 76.1 | 0.911 |
| | *Geometry* | *Hausdorff* [133] | 54.4 | 29.3 | 1.343 | 5.3 | 1.404 | 22.1 | 1.369 |
| | *Geometry* | *Hausdorff PSNR* [133] | 53.3 | 73.8 | 0.948 | 76 | 0.912 | 61 | 1.113 |
| | *Color + Geometry* | *Y-MSE* [104] | 70.3 | 59.3 | 1.131 | 70.3 | 0.998 | 52.4 | 1.196 |
| | *Color + Geometry* | *Y-PSNR* [104] | 70.3 | 71.3 | 0.985 | 71.4 | 0.984 | 71.3 | 0.985 |
| *Point-based (Po2Pl)* | *Geometry* | *D2* [104] | 90.6 | 76.2 | 0.910 | 87.9 | 0.669 | 69.9 | 1.004 |
| | *Geometry* | *D2-PSNR* [104] | 87.3 | 83.2 | 0.779 | 83.4 | 0.774 | 77.3 | 0.891 |
| | *Geometry* | *Hausdorff* [133] | 66.3 | 45.3 | 1.252 | 78.4 | 0.871 | 28.5 | 1.346 |
| | *Geometry* | *Hausdorff PSNR* [133] | 65.3 | 72.7 | 0.964 | 68.7 | 1.02 | 65.7 | 1.059 |
| *Point-based (Pl2Pl)* | *Geometry* | *MSE* [109] | 55.0 | 64.7 | 1.071 | 69 | 1.016 | 64.6 | 1.072 |
| *Feature-based* | *Geometry* | *PCQM* [134] | - | - | - | - | - | - | - |
| | *Geometry* | $d_{gc}$ [136] | 93.9 | 81.7 | 0.811 | 92.5 | 0.533 | 73.0 | 0.960 |
| | *Color + Geometry* | $H^{Y}_{L2}$ [136] | 91.9 | 83.4 | 0.776 | 87.9 | 0.669 | 75.5 | 0.922 |
| | *Color + Geometry* | $PCM_{RR}$*(MCCV)* [137] | 91.0 | 86.7 | 0.699 | 89.2 | 0.636 | 79.0 | 0.862 |
| | *Geometry, Color* | *Point SSIM* [135] | 92.9 | <u>94.4</u> | <u>0.465</u> | <u>94.4</u> | <u>0.462</u> | 92.8 | 0.522 |
| | *Color + Geometry* | *Geotex* [138] | - | - | - | - | - | - | - |
| *Projection-based* | *Color + Geometry* | *Y-MS-SSIM* [121] | 50.1 | 74.4 | 0.938 | 75.3 | 0.924 | 72.5 | 0.967 |
| | *Color + Geometry* | *Y-VIFP* [121] | 79.2 | 75.0 | 0.929 | 75.0 | 0.929 | 74.9 | 0.93 |
| | *Color + Geometry* | *Y-PSNR* [126] | 35.7 | 46.6 | 1.242 | 47.7 | 1.234 | 43.7 | 1.263 |
| *Point-based (Po2Po)* | *Geometry* | *GH 98% PSNR (Ch. 6)* | 89.9 | 88.6 | 0.651 | 88.5 | 0.653 | 82.7 | 0.790 |
| | *Geometry* | *RA-PSNR* ($APD_{10}$) *(Ch. 7)* | 91.8 | 90.0 | 0.611 | 91.0 | 0.584 | 81.1 | 0.821 |
| *Point-based (Po2Pl)* | *Geometry* | *GH 98% PSNR (Ch. 6)* | 91.0 | 87.8 | 0.672 | 87.5 | 0.68 | 82.3 | 0.798 |
| | *Geometry* | *RA-PSNR* ($APD_{10}$) *(Ch. 7)* | 91.5 | 90.3 | 0.605 | 90.3 | 0.604 | 82.1 | 0.803 |
| *Point-based (P2D)* | *Geometry* | *MMD (Ch. 9)* | 90.3 | 87.8 | 0.672 | 88.8 | 0.647 | 82.2 | 0.800 |
| | *Geometry* | *MMD-PSNR (Ch. 9)* | 90.3 | 88.3 | 0.658 | 88.7 | 0.648 | 88.3 | 0.659 |
| *Projection-based* | *Color + Geometry* | *ProjQm-FSIM (Ch. 10)* | 92.1 | 89.4 | 0.629 | 90.3 | 0.604 | 86.7 | 0.699 |
| | *Color + Geometry* | *ProjQm-VSI (Ch. 10)* | 90.1 | 86.1 | 0.715 | 88.1 | 0.664 | 81.8 | 0.807 |
| | *Color + Geometry* | *ProjQm-LPIPS (Ch. 10)* | <u>94.2</u> | 93.4 | 0.500 | 93.5 | 0.497 | <u>93.4</u> | <u>0.502</u> |
| | *Color + Geometry* | *ProjQm-DISTS (Ch. 10)* | **96.0** | **95.8** | **0.404** | **95.8** | **0.402** | **94.4** | **0.461** |

#### *9.3.2.5 Proposed Projection-based Quality Metric Ablation Study*

The proposed projection-based quality metric has many modules which have different impact on the overall performance. To assess individually each module of the proposed metric, different ablation studies are run on the entire dataset. More precisely, the quality metric is run for all stimuli included in the dataset, each time turning off one of the key modules while keeping the others: recoloring, cropping, and padding. Table 9.8 shows the PLCC and SROCC results for this ablation study. The non-linear fitting is not applied to the objective scores to eliminate any possible performance oscillations which could have an impact on the results and conclusions. The following conclusions can be derived:

- **Recoloring**: The correlation performance results show the importance of recoloring for the proposed projection-based quality metric. The absence of recoloring for VSI and FSIM leads to losses of 10.3 to 8.6 for SROCC and 3.6 to 2.4 for PLCC. The performance loss after removing recoloring for DISTS and LPIPS is lower mainly because these two recent 2D quality metric are robust to geometry distortions and transformations. However, the performance gain of using recoloring for these two projection-based quality metrics is respectively, 1.5 and 2.5 SROCC and 2.4 and 2.2 PLCC.
- **Cropping**: The cropping operation improves the performance of a projection-based metric significantly by removing background pixels around objects that are in common between reference and degraded PCs. The background area works as a distractor for the 2D quality assessment metric and decreases its prediction power, even when background pixels are padded. The performance gains by using cropping module are up to 11.4 for SROCC and 14.0 for PLCC.
- **Padding**: The padding operation also improves the performance of the proposed metric. Most of the excessive background is removed by cropping but still some area around the object remains. Also, background pixels among objects which are the result of sparce areas are filled by padding. These background pixels inside objects results in comparison of background pixels with point cloud pixels due to different number of decoded points.

Note that all of these performance gains for these modules may be significantly more using 2D image metrics with lower performance.

**Table 9.8:** Ablations studies of the proposed metric for 4 best 2D metrics on used dataset. Correlation results are obtained without fitting objective scores to MOS.

| | | **Final Proposed Metric** | **Removed Module** | | |
|---|---|---|---|---|---|
| | | | **Recoloring** | **Cropping** | **Padding** |
| ProjQM-DISTS | *PLCC* | 84.0 | 81.6 | 72.4 | 83.3 |
| | *SROCC* | 95.6 | 94.9 | 86.7 | 91.1 |
| ProjQM-LPIPS | *PLCC* | 88.0 | 85.8 | 73.9 | 76.5 |
| | *SROCC* | 93.2 | 90.7 | 81.8 | 80.8 |
| ProjQM-FSIM | *PLCC* | 74.3 | 71.9 | 63.8 | 73.6 |
| | *SROCC* | 90.1 | 81.5 | 83.7 | 89.6 |
| ProjQM-VSI | *PLCC* | 68.0 | 64.4 | 58.4 | 63.7 |
| | *SROCC* | 87.6 | 77.3 | 83.9 | 83.4 |

## 9.4 **Final Remarks**

This chapter proposes a novel projection-based point cloud quality metric that compares the point clouds in the 2D domain for the same geometry level (reference and degraded geometry). The projection-based quality metric obtains 2D project maps for six views of the point cloud which are processed (padding and cropping) before any 2D quality assessment. The performance results achieved confirm a very high performance, much better than other available state-of-the-art point cloud quality metrics. Other experimental results show, which 2D quality metric leads to the highest overall performance and which one of the modules of the proposed quality metric has a higher impact in the overall performance (ablation study). In this Chapter, the objective-subjective correlation results show that proposed metric outperforms all previous metrics proposed by the author during the development of this Thesis.

This work is going to be submitted soon.

**A. Javaheri,** C. Brites, F. Pereira, J. Ascenso, “Projection-based Point Cloud Quality Assessment Metric,” to be submitted.

# Chapter 10

# 10. Summary and Future Work

## 10.1 Conclusions

This Thesis has addressed key challenges in the emerging field of point cloud quality assessment with a special focus on the quality assessment of decoded point clouds. The subjective quality assessment studies reported in this Thesis have started with the quality assessment of simple point clouds, without attributes, decoded with the only coding scheme available at the time to move after to more complex contents with color, decoded with emerging point cloud coding standards.

Several subjective quality assessment studies have been performed to study and benchmark the impact of different factors on the perceived quality of point clouds e.g., rendering, coding, denoising etc. These studies were used later to evaluate the objective-subjective correlation performance of the proposed objective quality assessment metrics. In this context, several point cloud objective quality metrics have been proposed in this Thesis, trying to consider the key factor in the perceptual quality of decoded point clouds.

The summary and conclusions of this Thesis are, naturally, related to the two major parts of this Thesis: i) Subjective quality assessment; and ii) Objective quality assessment, as follows:

**Part II. Subjective Quality Assessment**

- **Impact of Denoising:** A passive subjective assessment methodology has been proposed to assess the quality of point cloud geometry denoising algorithms. No color attribute was used in this subjective study. To perform this study, classical image denoising solutions, e.g., using Tikhonov and Total Variation regularization functions, were adapted for point clouds, inspired by the emerging graph-based signal processing field to either remove erroneous points or improve its positioning. Experimental results showed that denoising with a total variation regularization function improves the quality of point clouds slightly more than with a Tikhonov

regularization function. The objective-subjective correlation performance of the two types of distance metrics available at that time were compared to conclude that Po2Pl metrics perform better than Po2Po metrics in evaluating the quality of denoised point clouds. However, Po2Po and Po2Pl Hausdorff distances showed very poor performance to evaluate denoised point clouds.

- **Impact of Coding:** A passive subjective assessment methodology has been proposed to assess the quality of decoded point clouds. Decoded point clouds with their reference color were used in the subjective tests. To perform this study, the PCL octree-based codec and a simple transform-based codec were used. The objective-subjective correlation performance of the distance-based metrics in the literature at that time, notably Po2Po/Po2Pl MSE and Hausdorff, were analysed to conclude that Po2Pl Hausdorff outperformed other quality metrics for point clouds decoded with both codecs.
- **Impact of Rendering:** A subjective study was performed in three sessions to study the impact of frequently used rendering approaches on the perceptual quality of decoded point clouds. This was the first time that the rendering and coding processes, which play a major role on the final perceived quality, were jointly evaluated. It was also the first time that the emerging MPEG point cloud coding standards, G-PCC and V-PCC, were used in a subjective quality assessment test. Experimental results have shown that, in general, the visibility of compression artifacts is different for different rendering scenarios, e.g., color attributes mask the geometry distortions. However, rendering affects the distortions associated to different coding approaches differently, e.g., PCL distortions are mostly visible in all rendering approaches while color highly masks errors associated to V-PCC. The objective-subjective correlation performance of nine of the most adopted state-of-the-art quality metrics e.g., D1-PSNR and D2-PSNR, was also compared for different rendering approaches. It was shown that the perceived quality not only depends on the point cloud artifacts introduced by some processing step (in this case, coding) but also on the rendering approach.

**Part III. Objective Quality Assessment**

- **GH-PSNR:** A novel point cloud geometry quality assessment metric was proposed based on the generalized Hausdorff distance. The analysis of the error histograms associated to the G-PCC and V-PCC codecs has shown that there are a few points with very large error magnitude comparing to rest of the points. These errors are not perceptually very visible because they are outnumbered and scattered all over the point cloud. The proposed PSNR-based metric uses the generalized Hausdorff distance with different percentages to filter these errors out. Experimental results have shown that the proposed metric outperforms the D1-PSNR and D2-PSNR metrics.
- **RA-PSNR:** A novel point cloud geometry quality assessment metric was proposed by exploiting the intrinsic point cloud characteristics and the rendering process that must occur before visualization. The well-known D1-PSNR and D2-PSNR metrics were improved by

including a normalization factor that accounts for changes in the intrinsic point cloud resolution (after rendering), as well as the point cloud precision. Experimental results have shown that the proposed quality metrics outperform the MPEG PSNR-based quality metrics significantly.

- **P2D PCQM:** A novel point cloud geometry quality assessment metric was proposed which establishes a new type of correspondence between a point and a distribution of points instead of the more usual point-to-point correspondence. When the decoded number of points is larger than the reference number of points in the reference point cloud, the point-to-point correspondence for the newly added points is not accurate. The quality metric extracts the statistics of a distribution of points and considers the underlying surface more effectively than available point-based quality metrics. Experimental results have shown a significant improvement comparing to the state-of-the-art quality metrics.
- **ProjQM:** A novel projection-based point cloud quality metric was proposed which addresses the limitations of the state-of-the-art projection-based metrics, notably achieving a better objective-subjective correlation performance. The underlying idea is that even a slight misalignment between the geometry of the reference and degraded point clouds may not be perceptually visible but may result in comparing non-corresponding pixels after the projection. The proposed quality metric exploits the recoloring process to align the geometries. It also crops the background and fills the holes inside the projected point cloud in order to discard common parts that reduce the metric prediction power, notably by preventing to compare a projected point with a background pixel. Experimental results have shown that by adopting the suitable 2D metric, the proposed quality metric became the best performing quality metric in the point cloud quality assessment.

In Chapter 9, all proposed objective quality metrics were compared using MOS scores and stimuli from a subjective quality assessment study with the M-PCCD dataset [121]. Table 10.1 shows the obtained quality metrics ranking in terms of objective-subjective correlation performance for the G-PCC and V-PCC codecs individually and both together. The best proposed metrics for all codecs together is projection-based metric proposed in Chapter 9. RA-PSNR (Chapter 7), P2D MMD (Chapter 8) and GH-PSNR (Chapter 6) are in the next rank.

**Table 10.1:** Ranking of the proposed objective quality metrics on their SROCC objective-subjective correlation performance for M-PCCD dataset.

| Rank | G-PCC | | | V-PCC | | | All | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Quality Metric** | **SROCC** | **PLCC** | **Quality Metric** | **SROCC** | **PLCC** | **Quality Metric** | **SROCC** | **PLCC** |
| **1** | ProjQm-DISTS | 96.0 | 95.8 | GAI-PCQM (DISTS) | 85.3 | 86.4 | ProjQm-DISTS | 95.6 | 94.7 |
| **2** | Po2Po RA-PSNR ($APD_{10}$) | 91.8 | 91.0 | Po2Pl RA-PSNR ($APD_{10}$) | 76.9 | 79.9 | Po2Po RA-PSNR ($APD_{10}$) | 90.2 | 88.8 |
| **3** | Po2Pl GH-PSNR 98% | 91.0 | 87.5 | Po2Pl GH-PSNR 98% | 71.2 | 75.1 | P2D MMD PSNR | 88.7 | 86.9 |
| **4** | P2D MMD PSNR | 90.3 | 88.3 | P2D MMD PSNR | 69.0 | 71.9 | Po2Pl GH-PSNR 98% | 87.9 | 84.3 |

## 10.2 Future Work

The research conducted in this Thesis has resulted in several subjective quality assessment studies addressing different problems in point cloud quality as well as in four novel objective quality metrics which have shown promising performance for point cloud quality assessment. However, there is still plenty of room to improve the current point cloud objective quality metrics. There are also many impacting factors that are not yet considered in point cloud subjective assessment.

In this context, the following topics may be considered for future research:

- Using other visualization devices, such VR glasses as well as 2D displays, in point cloud subjective quality assessment test to assess their impact on the user experience, notably the different levels of interactivity they may offer.
- Using more rendering approaches for point cloud subjective quality studies, notably spanning the requirements of more application scenarios.
- Exploiting visual attention in virtual or augment reality scenarios, to see how people consume 3D point clouds when they are free to interact with them.
- Exploring deep learning-based metrics considering the promising results of feature-based metrics in point cloud quality assessment and deep learning methods in other computer science domains.
- Exploiting point cloud statistics to propose a no-reference point cloud quality assessment.

- Evaluating dynamic point clouds using either subjective studies or by proposing objective quality metrics.

Adopting a feature-based approach for the objective quality metrics which is already showing promising results in the literature. Feature of reference and degraded can be learned through a feature learning neural network and compared together. Deep neural networks have been used for image quality assessment [145] [146] and point cloud geometry similarity in registration validation used in robotic application [148]. Since features extracted by a deep neural network have never been used for point cloud quality assessment, this seems an interesting approach to pursue.

# References


[1] C. Tulvan, R. Mekuria, Z. Li and S. Laserre, "Use cases for Point Cloud Compression (PCC)," in *ISO/IEC JTC1/SC29/WG11 MPEG2015/ N16331*, Geneva, Switzerland, June, 2016.

[2] 3DG Group, "Common Test Conditions for Point Cloud Compression," ISO/IEC JTC1/SC29/WG11 Doc. N18474, Geneva, Switzerland, March, 2019.

[3] 3DG Group, "Text of ISO/IEC CD 23090-9 Geometry-based Point Cloud Compression," ISO/IEC JTC1/SC29/WG11 Doc. N18478, Geneva, Switzerland, July, 2019.

[4] 3DG Group, "Text of ISO/IEC CD 23090-5: Video-based Point Cloud Compression," ISO/IEC JTC1/SC29/WG11 Doc. N18030, Macau, China, October, 2018.

[5] E. Alexiou, E. Upenik and T. Ebrahimi, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *Workshop on Multimedia Signal Processing*, Luton, UK, October, 2017.

[6] E. Alexiou and T. Ebrahimi, "On subjective and objective quality evaluation of point cloud geometry," in *International Conference on Quality of Multimedia Experience*, Erfurt, Germany, May, 2017.

[7] E. Alexiou and T. Ebrahimi, "On the performance of metrics to predict quality in point cloud representations," in *Applications of Digital Image Processing XL (SPIE 10396)*, San Diego, CA, USA, September, 2017.

[8] E. Alexiou, T. Ebrahimi, M. V. Bernardo, M. Pereira, A. Pinheiro, L. A. da Silva Cruz, C. Duarte, L. G. Dmitrovic, E. Dumic, D. Matkovics and A. Skodras, "Point cloud subjective evaluation methodology based on 2D rendering," in *International Conference on Quality of Multimedia Experience*, Sardinia, Italy, May, 2018.

[9] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Subjective and objective quality evaluation of 3D point cloud denoising algorithms," in *International Conference on Multimedia & Expo Workshops*, Hong Kong, July, 2017.

[10] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Subjective and objective quality evaluation of compressed point clouds," in *Workshop on Multimedia Signal Processing*, Luton, UK, October, 2017.

[11] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Point Cloud Rendering after Coding: Impacts on Subjective and Objective Quality," *Transactions on Multimedia*, 2020.

[12] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "A generalized Hausdorff distance based quality metric for point cloud geometry," in *International Conference on Quality of Multimedia Experience*, Athlone, Ireland, May, 2020.

[13] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Improving PSNR-based quality metrics performance for point cloud geometry," in *International Conference on Image Processing*, Abu Dhai, UAE, October, 2020..

[14] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Mahalanobis Based Point to Distribution Metric for Point Cloud Geometry Quality Evaluation," *Signal Processing Letters*, vol. 27, pp. 1350-1354, 2020.

[15] R. B. Rusu and A. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics and Automation*, Shanghai, China, May, 2011.

[16] F. Pereira, A. Dricot, J. Ascenso and C. Brites, "Point cloud coding: A privileged view driven by a classification taxonomy," *Signal Processing: Image Communication*, vol. 85, no. 115862, July, 2020.

[17] J. Kammerl, N. Blodow, R. B. Rusu and S. Gedikli, "Real-time compression of point cloud streams," in *International Conference on Robotics and Automation*, Saint Paul, MN, USA, May, 2012.

[18] G. Nigel and N. Martin, "Range encoding: an algorithm for removing redundancy from a digitized message," in *Video and Data Recording Conference*, Southampton, UK, July, 1979.

[19] 3DG Group, "G-PCC codec description v7," in *ISO/IEC MPEG N19331*, Online, April. 2020.

[20] R. Queiroz and P. Chou, "Compression of 3D Point Clouds Using a Region-Adaptive Hierarchical Transform," *IEEE Transaction on Image Processing,* vol. 25, no. 8, pp. 3947 - 3956, August, 2016.

[21] "V-PCC codec description," in *ISO/IEC JTC 1/SC 29/WG 7 N00012*, Virtual, October, 2020.

[22] G. J. Sullivan, J. R. Ohm, W. Han and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 22, no. 12, pp. 1649-1668, September, 2012.

[23] M. Pharr, W. Jakob and G. Humphreys, Physically based rendering: from theory to implementation,, Morgan Kaufmann, 2016.

[24] T. Akenine-Moller, E. Haines and N. Hoffman, Real-time rendering, vol. 3rd edition, AK Peters/CRC Press, 2018.

[25] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Subjective and Objective Quality Evaluation of Compressed Point Clouds," in *Workshop on Multimedia Signal Processing (MMSP),* Luton, UK, October 2017.

[26] V. S. Ramachandran, "Perception of Shape from Shading," *Nature,* vol. 331, no. 6152, pp. 163-166, January 1988.

[27] M. Kazhdan, M. Bolitho and H. Hoppe, "Poisson Surface Reconstruction," in *Eurographics Symp. on Geometry Processing*, Sardinia, Italy, June 2006.

[28] A. Khatamian and H. R. Arabnia, "Survey on 3D Surface Reconstruction," *Journal of Information Processing Systems,* vol. 12, no. 3, pp. 338-357, September 2016.

[29] R. Zhang, P. S. Tsai, J. E. Cryer and M. Shah, "Shape-from-shading: a survey," *Transactions on Pattern Analysis and Machine Intelligence,* vol. 2, no. 8, pp. 690-706, August 1999.

[30] D. F. Fouhey, A. Gupta and A. Zisserman, "From Images to 3D Shape Attributes," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 41, no. 1, pp. 93-106, December 2017.

[31] ITU-R Recommendation BT.500-13, "Methodology for the Subjective Assessment of the Quality of Television Pictures," in *International Telecommunications Union*, Geneva, Switzerland, 2012.

[32] Tutorial, ITU-T, "Objective Perceptual Assessment of Video Quality: Full Reference Television," in *ITU-T Telecommunication Standardization Bureau*, 2005.

[33] ITU-R Recommendation P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," Geneva, Switzerland, July 2012.

[34] J. Zhang, W. Huang, X. Zhu and J.-N. Hwang, "A subjective quality evaluation for 3D point cloud models," in *International Conerance. on Audio, Language and Image Processing*, Shanghai, China, July, 2014.

[35] R. Mekuria, K. Blom and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828 - 842, April, 2017.

[36] R. Mekuria, S. Laserre and C. Tulvan, "Performance assessment of point cloud compression," in *Visual Communications and Image Processing (VCIP)*, St. Petersburg, FL, USA, December, 2017.

[37] E. Alexiou and T. Ebrahimi, "On the performance of metrics to predict quality in point cloud representations," in *Applications of Digital Image Processing XL (SPIE 10396)*, San Diego, CA, USA, September, 2017.

[38] E. Alexiou and T. Ebrahimi, "Impact of visualization strategy for subjective quality assessment of point clouds," in *International Conference on Multimedia & Expo Workshops*, San Diego, CA, USA, July 2018.

[39] E. Dumic, F. Battisti, m. Carli and L. A. da Silva, "Point Cloud Visualization Methods: a Study on Subjective Preferences," in *European Signal Processing Conference (EUSIPCO)*, Amsterdam, Netherlands, January, 2021.

[40] JPEG Committee, "JPEG Pleno database - University of Sao Paulo point clouds," [Online]. Available: http://plenodb.jpeg.org/.

[41] E. Alexiou, A. M. G. Pinheiro, C. Duarte, D. Matković, E. Dumić and L. A. da Silva Cruz, "Point cloud subjective evaluation methodology based on reconstructed surfaces," in *Applications of Digital Image Processing XL (SPIE 10752)*, San Diego, CA, USA, September, 2018.

[42] E. Alexiou and T. Ebrahimi, "Benchmarking of objective quality metrics for colorless point clouds," in *Picture Coding Symposium*, San Francisco, CA, USA, June 2018.

[43] K. Christaki, E. Christakis, P. Drakoulis, A. Doumanoglou, N. Zioulis, D. Zarpalas and P. Daras, "Subjective visual quality assessment of immersive 3D media compressed by open-source static 3D mesh codecs," in *International Conference on Multimedia Modeling*, Thessaloniki, Greece, January, 2019.

[44] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," in *Applications of Digital Image Processing XLI*, San Diego, CA, USA, September, 2018.

[45] E. Alexiou and T. Ebrahimi, "Exploiting user interactivity in quality assessment of point cloud imaging," in *IEEE International Conference on Quality of Multimedia Experience (QoMEX)*, Berlin, Germany, June, 2019.

[46] E. Dumic, C. R. Duarte and L. A. da Silva Cruz, "Subjective evaluation and objective measures for point clouds — State-of-the-art," in *International Colloquium on Smart Grid Metrology*, Split, Croatia, May 2018.

[47] E. Alexiou, I. Viola, T. . M. Borges, T. A. Fonseca and R. L. de Queiroz, "A comprehensive study of the rate-distortion performance in MPEG point cloud compression," *APSIPA Transactions on Signal and Information Processing*, vol. 8, no. 27, pp. 1-27, November, 2019.

[48] E. Zerman, P. Gao, C. Ozcinar and A. Smolic, "Subjective and objective quality assessment for volumetric video compression," *Image Quality and System Performance*, vol. XVI, pp. 323-1-323-7, January, 2019.

[49] E. Zerman, C. Ozcinar, P. Gao and A. Smolic, "Textured Mesh vs Coloured Point Cloud: A Subjective Study for Volumetric Video Compression," in *International Conference on Quality of Multimedia Experience (QoMEX)*, Athlone, Ireland, May, 2020.

[50] "Draco," Google, [Online]. Available: https://google.github.io/draco/. [Accessed 25 01 2020].

[51] E. d'Eon, B. Harrison, T. Myers and P. A. Chou, "8i voxelized full bodies, version 2 – A voxelized point cloud dataset," in *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document m40059/M74006*, Geneva, Switzerland, January, 2017.

[52] X. Wu, Y. Zhang, F. Chun-ling, J. Hou and S. Kwong, "Subjective Quality Study and Database of Compressed Point Clouds with 6DoF Head-mounted Display," *arXiv: Image and Video Processing*, 2020.

[53] H. Su, Z. Duanmu, W. Liu, Q. Liu and Z. Wang, "Perceptual Quality Assessment of 3d Point Clouds," in *International Conference on Image Processing (ICIP)*, Taipei, Taiwan, September, 2019.

[54] J. van der Hooft, M. T. Vega, C. Timmerer, A. C. Begen, F. De Turck and R. Schatz, "Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming," in *International Conference on Quality of Multimedia Experience (QoMEX)*, Athlone, Ireland, May, 2020.

[55] S. Subramanyam, J. Li, I. Viola and P. Cesar, "Comparing the Quality of Highly Realistic Digital Humans in 3DoF and 6DoF: A Volumetric Video Case Study," in *Conference on Virtual Reality and 3D User Interfaces (VR)*, Atlanta, GA, USA, March, 2020.

[56] S. Perry, H. P. Cong, L. A. da Silva Cruz, J. Prazeres, M. Pereira, A. Pinheiro, E. Dumic, E. Alexiou and T. Ebrahimi, "Quality Evaluation Of Static Point Clouds Encoded Using MPEG Codecs," in *International Conference on Image Processing (ICIP)*, Abu Dhabi, UAE, October, 2020.

[57] R. Mekuria, Z. Li, C. Tulvan and P. Chou, "Evaluation criteria for PCC (Point Cloud Compression)," in *ISO/IEC MPEG Doc. N16332*, Geneva, Switzerland, June, 2016.

[58] D. Tian, H. Ochimizu, C. Feng, R. Cohen and A. Vetro, "Geometric distortion metrics for point cloud compression," in *IEEE International Conference on Image Processing (ICIP)*, Beijing, China, September, 2017.

[59] E. Alexiou and T. Ebrahimi, "Point Cloud Quality Assessment Metric based on Angular Similarity," in *IEEE International Conference on Multimedia & Expo Workshops*, San Diego, CA, USA, July, 2018.

[60] P. Cignoni, C. Rochinni and R. Scopigno, "Metro: measuring errors on simplified surfaces," *Computer Graphics Forum*, vol. 17, no. 2, pp. 167-174, January, 1998.

[61] G. Meynet, J. Digne and G. Lavoué , "PC-MSDM: A quality metric for 3D point clouds," in *International Conference on Quality of Multimedia Experience*, Berlin, Germany, June, 2019.

[62] I. Viola, S. Subramanyam and P. Cesar, "A color-based objective quality metric for point cloud contents," in *IEEE International Conference on Quality of Multimedia Experience*, Athlone, Ireland, May, 2020.

[63] R. Diniz, P. G. Freitas and M. Farias, "Towards a point cloud quality assessment model using local binary patterns," in *IEEE International Conference on Quality of Multimedia Experience*, Athlone, Ireland, May, 2020.

[64] I. Vadja, "On the F-Divergence and Singularity of Probability Measures," *Periodica Mathematica Hungarica,* vol. 2, no. 1-4, p. 223–234, March 1972.

[65] R. Diniz, P. G. Freitas and M. C. Q. Farias, "Multi-Distance Point Cloud Quality Assessment," in *IEEE International Conference on Image Processing (ICIP),* Abu Dhabi, UAE, October, 2020.

[66] R. Diniz, P. G. Freitas and M. C. Q. Farias, "Local Luminance Patterns for Point Cloud Quality Assessment," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Tampere, Finland, September, 2020.

[67] G. Meynet, Y. Nehmé, J. Digne and G. Lavoué, "PCQM: A Full-Reference Quality Metric for Colored 3D Point Clouds," in *IEEE International Conference on Quality of Multimedia Experience (QoMEX),* Athlone, Ireland, May, 2020.

[68] I. Viola and P. Cesar, "A Reduced Reference Metric for Visual Quality Evaluation of Point Cloud Contents," *IEEE Signal Processing Letters,* vol. 27, pp. 1660-1664, 2020.

[69] E. Alexiou and T. Ebrahimi, "Towards a Point Cloud Structural Similarity Metric," in *IEEE International Conference on Multimedia & Expo Workshops (ICMEW),* London, UK, July, 2020.

[70] R. L. de Queiroz and P. A. Chou, "Motion-Compensated Compression of Dynamic Voxelized Point Clouds," *IEEE Transactions on Image Processing,* vol. 26, no. 8, pp. 3886-3895, 2017.

[71] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang and J. Sun, "Predicting the Perceptual Quality of Point Cloud: A 3D-to-2D Projection-Based Exploration," *IEEE Transactions on Multimedia (Early Access),* 2020.

[72] K. Wolff, C. Kim, H. Zimmer, C. Schroers, M. Botsch, O. Sorkine-Hornung and A. Sorkine-Hornung, "Point Cloud Noise and Outlier Removal for Image-Based 3D Reconstruction," in *IEEE International Conference on 3D Vision (3DV),* Stanford, CA, USA, October, 2016.

[73] M. Berger, A. Tagliasacchi, L. M. Seversky, P. Alliez, G. Guennebaud, J. A. Levine, A. Sharf and C. T. Silva, "A Survey of Surface Reconstruction from Point Clouds," *Computer Graphics Forum,* vol. 36, no. 1, p. 301–329, January, 2017.

[74] Y. Schoenenberger, J. Paratte and P. Vanderg, "Graph-based Denoising for Time-Varying Point Clouds," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, Lisbon, Portugal, July, 2015.

[75] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha and M. Beetz, "Towards 3D Point Cloud Based Object Maps for Household Environments," *Robotics and Autonomous Systems,* vol. 56, no. 11, pp. 927-941, November, 2008.

[76] L. I. Rudin, S. Osher and E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D: Nonlinear Phenomena,* vol. 60, no. 1-4, pp. 259-268, November, 1992.

[77] D. I. Shuman, S. K. Narang and P. Frossard, "The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains," *IEEE Signal Processing Magazine,* vol. 30, no. 3, pp. 83-98, April, 2013.

[78] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning,* vol. 3, no. 1, pp. 1-122, January, 2011.

[79] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst and D. K. Hammond, "GSPBOX: A Toolbox for Signal Processing on Graphs," *arXiv preprint arXiv:1408.5781,* March, 2016.

[80] N. Perraudin, V. Kalofolias, D. Shuman and P. Vandergheynst, "UNLocBoX A Matlab Convex Optimization Toolbox Using Proximal Splitting Methods," *arXiv preprint arXiv:1402.0779,* December, 2016.

[81] M. Levoy, J. Gerth, B. Curless and K. Pull, "The Stanford 3D Scanning Repository," Stanford University, 2005. [Online]. Available: http://graphics.stanford.edu/data/3Dscanrep/.

[82] M. Kazhdan, M. Bolitho and H. Hoppe, "Poisson surface reconstruction," in *Eurographics Symp. on Geometry Processing*, Sardinia, Italy, June, 2006.

[83] D. Girardeau-Montaut, "Cloud Compare—3D Point Cloud and Mesh Processing Software," [Online]. Available: http://www.cloudcompare.org/.

[84] L. Goldmann, F. De Simone and T. Ebrahimi, "A Comprehensive Database and Subjective Evaluation Methodology for Quality of Experience in Stereoscopic Video,"," in *Three-Dimensional Image Processing (3DIP) and Application*, San Jose, CA, USA, February, 2010.

[85] R. Mekuria, K. Blom and P. Cesar, "Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-immersive Video," *Transactions on Circuits and Systems for Video Technology,* vol. 27, no. 4, pp. 828 - 842, April 2017.

[86] J. Kammerl, N. Blodow, R. B. Rusu and S. Gedikli, "Real-time Compression of Point Cloud Streams," in *International Conference on Robotics and Automation*, Saint Paul, MN, USA, May 2012.

[87] D. Thanou, P. A. Chou and P. Frossard, "Graph-based Compression of Dynamic 3D Point Cloud Sequences," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1765-1778, 2017.

[88] R. A. Cohen, D. Tian and A. Vetro, "Point Cloud Attribute Compression using 3-D Intra Prediction and Shape-adaptive Transforms," in *IEEE Data Compression Conference (DCC)*, Snowbird, UT, USA, March 2016.

[89] R. B. Rusu and A. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics and Automation*, Shanghai, China, May 2011.

[90] D. I. Shuman, S. K. Narang and P. Frossard, "The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains," *IEEE Signal Processing Magazine,* vol. 30, no. 3, pp. 83-98, April 2013.

[91] "MPEG Point Cloud Datasets," [Online]. Available: http://mpegfs.int-evry.fr/MPEG/PCC/DataSets/pointCloud/CfP.

[92] G. Guede, J. Ricard, S. Lasserre and J. Llach, "Technicolor Point Cloud Renderer," in *ISO/IEC MPEG M40229*, Hobart, Australia, April 2017.

[93] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Subjective and Objective Quality Evaluation of 3D Point Cloud Denoising Algorithms," in *International Conference on Multimedia & Expo Workshops*, Hong Kong, July 2017.

[94] M. Schütz and M. Wimmer, "High-quality point-based rendering using fast single-pass interpolation," in *IEEE Digital Heritage*, Granada, Spain, September 2015.

[95] P. Rosenthal and L. Linsen, "Image-space point cloud rendering," in *International Conference on Computer Graphics*, Istanbul, Turkey, June 2008.

[96] E. Alexiou, E. Upenik and T. Ebrahimi, "Towards Subjective Quality Assessment of Point Cloud Imaging in Augmented Reality," in *Workshop on Multimedia Signal Processing*, Luton, UK, October 2017.

[97] E. Alexiou, T. Ebrahimi, M. V. Bernardo, M. Pereira, A. Pinheiro, L. A. da Silva Cruz, C. Duarte, L. G. Dmitrovic, E. Dumic, D. Matkovics and A. Skodras, "Point Cloud Subjective Evaluation Methodology based on 2D Rendering," in *International Conference on Quality of Multimedia Experience*, Sardinia, Italy, May 2018.

[98] E. Alexiou and T. Ebrahimi, "Impact of Visualization Strategy for Subjective Quality Assessment of Point Clouds," in *International Conference on Multimedia & Expo Workshops*, San Diego, CA, USA, July 2018.

[99] E. Alexiou, A. M. G. Pinheiro, C. Duarte, D. Matković, E. Dumić and L. A. da Silva Cruz, "Point Cloud Subjective Evaluation Methodology based on Reconstructed Surfaces," in *Applications of Digital Image Processing XL (SPIE 10752)*, San Diego, CA, USA, September 2018.

[100] E. Alexiou and T. Ebrahimi, "Benchmarking of Objective Quality Metrics for Colorless Point Clouds," in *Picture Coding Symposium*, San Francisco, CA, USA, June 2018.

[101] P. Cignoni, M. Callieri, M. Corsi, M. Dellepiane, F. Ganovelli and G. Ranzuglia, "Meshlab: an Open-Source Mesh Processing Tool," *Eurographics Italian Chapter Conference* , vol. 2008, pp. 129-136, July 2008.

[102] 3DG Group, "G-PCC codec description v7," in *ISO/IEC MPEG N19331*, Online, April 2020.

[103] 3DG Group, "Text of ISO/IEC CD 23090-5: Video-based Point Cloud Compression," ISO/IEC JTC1/SC29/WG11 Doc. N18030, Macau, China, October 2018.

[104] 3DG Group, "Common Test Conditions for Point Cloud Compression," ISO/IEC JTC1/SC29/WG11 Doc. N18474, Geneva, Switzerland, March 2019.

[105] A. Javaheri, "IST Rendering Point Cloud Dataset," 2019. [Online]. Available: https://github.com/AlirezaJav/IRPC-Dataset. [Accessed 23 April 2019].

[106] I. Viola and T. Ebrahimi, "An in-depth analysis of single-image subjective quality assessment of light field contents," in *International Conference on Quality of Multimedia Experience*, Berlin, Germany, June 2019.

[107] R. K. Mantiuk, A. Tomaszewska and R. Mantiuk, "Comparison of four subjective methods for image quality assessment," *Computer Graphics Forum*, vol. 31, no. 8, pp. 2478-2491, December 2012.

[108] Y. Nehmé, J. P. Farrugia, F. Dupont, P. Le Callet and G. Lavoué, "Comparison of subjective methods, with and without explicit reference, for quality assessment of 3D graphics," in *ACM Symposium on Applied Perception*, Barcelona, Spain, September 2019.

[109] E. Alexiou and T. Ebrahimi, "Point Cloud Quality Assessment Metric based on Angular Similarity," in *IEEE International Conference on Multimedia & Expo Workshops*, San Diego, CA, USA, July 2018.

[110] S. Athar and Z. Wang, "A comprehensive performance evaluation of image quality assessment algorithms," *IEEE Access*, vol. 7, pp. 140030-140070, September 2019.

[111] J. H. Kim and L. Choi, "Choosing the level of significance: a decision-theoretic approach," *Journal of Accounting, Finance and Business Studies*, November 2019.

[112] E. Alexiou and T. Ebrahimi, "On the Performance of Metrics to Predict Quality in Point Cloud Representations," in *Applications of Digital Image Processing XL (SPIE 10396)*, San Diego, CA, USA, September 2017.

[113] 3DG Group, "Text of ISO/IEC CD 23090-9 Geometry-based Point Cloud Compression," ISO/IEC JTC1/SC29/WG11 Doc. N18478, Geneva, Switzerland, July 2019.

[114] D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, September 1993.

[115] M. P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *International Conference on Pattern Recognition*, Jerusalem, Israel, October 1994.

[116] R. Mekuria, C. Tulvan and Z. Li, "Requirements for point cloud compression," in *ISO/IEC MPEG N16330*, Geneva, Switzerland., February 2016..

[117] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Point Cloud Rendering after Coding: Impacts on Subjective and Objective Quality," *Transactions on Multimedia*, November 2020.

[118] D. Tian, H. Ochimizu, C. Feng, R. Cohen and A. Vetro, "Geometric Distortion Metrics for Point Cloud Compression," in *IEEE International Conference on Image Processing (ICIP)*, Beijing, China, September 2017.

[119] C. Mineo, S. Gareth and P. R. Summan, "Novel algorithms for 3D surface point cloud boundary detection and edge reconstruction," *Journal of Computational Design and Engineering*, vol. 6, no. 1, pp. 81-91, January 2019.

[120] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, USA, November 2003.

[121] E. Alexiou, I. Viola, T. . M. Borges, T. A. Fonseca and R. L. de Queiroz, "A Comprehensive Study of the Rate-Distortion Performance in MPEG Point Cloud Compression," *APSIPA Transactions on Signal and Information Processing*, vol. 8, no. 27, pp. 1-27, November 2019.

[122] Y. Zhang, B. Du, L. Zhang and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1376-1389, March 2016.

[123] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207-244, February 2009.

[124] C. N. Haddad, "Cholesky factorization," in *Encyclopedia of Optimization*, 2nd ed., Boston, MA, USA, Springer, 2008, pp. 374-377.

[125] G. Meynet, J. Digne and G. Lavoué , "PC-MSDM: A Quality Metric for 3D Point Clouds," in *International Conference on Quality of Multimedia Experience*, Berlin, Germany, June 2019.

[126] R. L. de Queiroz and P. A. Chou, "Motion-Compensated Compression of Dynamic Voxelized Point Clouds," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3886-3895, May 2017.

[127] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz and T. Ebrahimi, "A Novel Methodology for Quality Assessment of Voxelized Point Clouds," in *Applications of Digital Image Processing XLI*, San Diego, CA, USA, September 2018.

[128] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang and J. Sun, "Predicting the Perceptual Quality of Point Cloud: A 3D-to-2D Projection-Based Exploration," *IEEE Transactions on Multimedia (Early Access)*, October 2020.

[129] M. Bertalmio, A. L. Bertozzi and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Kauai, HI, USA, December 2001.

[130] MMSPG, EPFL, "M-PCCD: MPEG Point Cloud Compression Dataset," [Online]. Available: https://www.epfl.ch/labs/mmspg/downloads/quality-assessment-for-point-cloud-compression/.

[131] JPEG Committee, "JPEG Pleno Database - University of Sao Paulo Point Clouds," [Online]. Available: http://plenodb.jpeg.org/.

[132] "Sketchfab," [Online]. Available: https://sketchfab.com/.

[133] R. Mekuria, Z. Li, C. Tulvan and P. Chou, "Evaluation Criteria for PCC (Point Cloud Compression)," in *ISO/IEC MPEG Doc. N16332*, Geneva, Switzerland, June 2016.

[134] G. Meynet, Y. Nehmé, J. Digne and G. Lavoué, "PCQM: A Full-Reference Quality Metric for Colored 3D Point Clouds," in *IEEE International Conference on Quality of Multimedia Experience (QoMEX)*, Athlone, Ireland, May 2020.

[135] E. Alexiou and T. Ebrahimi, "Towards a Point Cloud Structural Similarity Metric," in *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, London, UK, July 2020.

[136] I. Viola, S. Subramanyam and P. Cesar, "A Color-based Objective Quality Metric for Point Cloud Contents," in *IEEE International Conference on Quality of Multimedia Experience*, Athlone, Ireland, May 2020.

[137] I. Viola and P. Cesar, "A Reduced Reference Metric for Visual Quality Evaluation of Point Cloud Contents," *IEEE Signal Processing Letters*, vol. 27, pp. 1660-1664, September 2020.

[138] R. Diniz, P. G. Freitas and M. Farias, "Towards a Point Cloud Quality Assessment Model Using Local Binary Patterns," in *IEEE International Conference on Quality of Multimedia Experience*, Athlone, Ireland, May 2020.

[139] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti and M. Carli, "New full-reference quality metrics based on HVS," in *Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, Arizona, USA, January 2006.

[140] N. Ponomarenko, F. Silvestri, K. Egiaz, M. Carli, J. Astola and V. Lukin, "On between-Coefficient Contrast Masking," in *International Workshop on Video Processing and Quality Metrics* , Scottsdale, Arizona, USA, January 2007.

[141] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing,* vol. 13, no. 4, pp. 600-612, April 2004.

[142] H. R. Sheikh and A. C. Bovik, "Image Information and Visual Quality," *IEEE Transactions on Image Processing,* vol. 15, no. 2, pp. 430-444, January 2006.

[143] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: A Feature SIMilarity Index for Image Quality Assessment," *IEEE Transactions on Image Processing,* vol. 20, no. 8, pp. 2378 - 2386, January 2011.

[144] L. Zhang, Y. Shen and H. Li, "VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment," *IEEE Transactions on Image Processing,* vol. 23, no. 10, pp. 4270 - 4281, August 2014.

[145] R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June, 2018.

[146] K. Ding, K. Ma, S. Wang and E. P. Simoncell, "Image Quality Assessment: Unifying Structure and Texture Similarity," *arXiv preprint arXiv:2004.07728,* April 2020.

[147] R. Reisenhofer, S. Bosse, G. Kutynio and T. Wiegand, "A Haar Wavelet-Based Perceptual Similarity Index for Image Quality Assessment," *Signal Processing: Image Communication,* vol. 61, pp. 33-43, February 2018.

[148] J. Mazumder, M. Zand, S. Ziauddin and M. Greenspan, "ValidNet: A Deep Learning Network for Validation of Surface Registration," in *International Conference on Computer Vision Theory and Applications*, Valleta, Malta, February 2020.

[149] "vsenseVVDB2," [Online]. Available: https://v-sense.scss.tcd.ie/research/6dof/quality-assessment-for-fvv-compression/.

[150] T. Akenine-Moller, E. Haines and N. Hoffman, Real-time Rendering, vol. 3rd edition, AK Peters/CRC Press, 2018.

[151] E. Alexiou and T. Ebrahimi, "Exploiting User Interactivity in Quality Assessment of Point Cloud Imaging," in *IEEE International Conference on Quality of Multimedia Experience (QoMEX)*, Berlin, Germany, June 2019.

[152] E. Alexiou and T. Ebrahimi, "On Subjective and Objective Quality Evaluation of Point Cloud Geometry," in *International Conference on Quality of Multimedia Experience*, Erfurt, Germany, May 2017.

[153] M. Berger, A. Tagliasacchi, L. M. Seversky, P. Alliez, G. Guennebaud, J. A. Levine, A. Sharf and C. T. Silva, "A Survey of Surface Reconstruction from Point Clouds," *Computer Graphics Forum,* vol. 36, no. 1, p. 301–329, January 2017.

[154] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, January 2011.

[155] K. Christaki, E. Christakis, P. Drakoulis, A. Doumanoglou, N. Zioulis, D. Zarpalas and P. Daras, "Subjective Visual Quality Assessment of Immersive 3D Media Compressed by Open-Source Static 3D Mesh Codecs," in *International Conference on Multimedia Modeling*, Thessaloniki, Greece, January 2019.

[156] P. Cignoni, C. Rochinni and R. Scopigno, "Metro: Measuring Errors on Simplified Surfaces," *Computer Graphics Forum*, vol. 17, no. 2, pp. 167-174, January 1998.

[157] E. d'Eon, B. Harrison, T. Myers and P. A. Chou, "8i Voxelized Full Bodies, Version 2 – A Voxelized Point Cloud Dataset," in *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document m40059/M74006*, Geneva, Switzerland, January 2017.

[158] R. Diniz, P. G. Freitas and M. C. Q. Farias, "Local Luminance Patterns for Point Cloud Quality Assessment," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Tampere, Finland, September 2020.

[159] R. Diniz, P. G. Freitas and M. C. Q. Farias, "Multi-Distance Point Cloud Quality Assessment," in *IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, UAE, October 2020.

[160] E. Dumic, F. Battisti, m. Carli and L. A. da Silva, "Point Cloud Visualization Methods: a Study on Subjective Preferences," in *European Signal Processing Conference (EUSIPCO)*, Amsterdam, Netherlands, January 2021.

[161] E. Dumic, C. R. Duarte and L. A. da Silva Cruz, "Subjective Evaluation and Objective Measures for Point Clouds — State-of-the-art," in *International Colloquium on Smart Grid Metrology*, Split, Croatia, May 2018.

[162] L. Goldmann, F. De Simone and T. Ebrahimi, "A Comprehensive Database and Subjective Evaluation Methodology for Quality of Experience in Stereoscopic Video,"," in *Three-Dimensional Image Processing (3DIP) and Application*, San Jose, CA, USA, February 2010.

[163] R. L. Graham and P. Hell, "On the History of the Minimum Spanning Tree Problem," *Annals of the History of Computing*, vol. 7, no. 1, pp. 43-57, January-March 1985.

[164] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "A Generalized Hausdorff Distance Based Quality Metric for Point Cloud Geometry," in *International Conference on Quality of Multimedia Experience*, Athlone, Ireland, May 2020.

[165] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Improving PSNR-based Quality Metrics Performance for Point Cloud Geometry," in *International Conference on Image Processing*, Abu Dhai, UAE, October 2020.

[166] A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Mahalanobis Based Point to Distribution Metric for Point Cloud Geometry Quality Evaluation," *Signal Processing Letters,* vol. 27, pp. 1350-1354, July 2020.

[167] R. Mekuria, S. Laserre and C. Tulvan, "Performance Assessment of Point Cloud Compression," in *Visual Communications and Image Processing (VCIP),* St. Petersburg, FL, USA, December 2017.

[168] G. Nigel and N. Martin, "Range Encoding: An Algorithm for Removing Redundancy from a Digitized Message," in *Video and Data Recording Conference*, Southampton, UK, July 1979.

[169] F. Pereira, A. Dricot, J. Ascenso and C. Brites, "Point Cloud Coding: A Privileged View Driven by a Classification Taxonomy," *Signal Processing: Image Communication,* vol. 85, no. 115862, July 2020.

[170] N. Perraudin, V. Kalofolias, D. Shuman and P. Vandergheynst, "UNLocBoX A Matlab Convex Optimization Toolbox Using Proximal Splitting Methods," *arXiv preprint arXiv:1402.0779,* December 2016.

[171] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst and D. K. Hammond, "GSPBOX: A Toolbox for Signal Processing on Graphs," *arXiv preprint arXiv:1408.5781,* March 2016.

[172] S. Perry, H. P. Cong, L. A. da Silva Cruz, J. Prazeres, M. Pereira, A. Pinheiro, E. Dumic, E. Alexiou and T. Ebrahimi, "Quality Evaluation Of Static Point Clouds Encoded Using MPEG Codecs," in *International Conference on Image Processing (ICIP),* Abu Dhabi, UAE, October 2020.

[173] M. Pharr, W. Jakob and G. Humphreys, Physically Based Rendering: from Theory to Implementation,, Morgan Kaufmann, 2016.

[174] R. Queiroz and P. Chou, "Compression of 3D Point Clouds Using a Region-Adaptive Hierarchical Transform," *IEEE Transaction on Image Processing,* vol. 25, no. 8, pp. 3947 - 3956, August 2016.

[175] L. I. Rudin, S. Osher and E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D: Nonlinear Phenomena,* vol. 60, no. 1-4, pp. 259-268, November 1992.

[176] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha and M. Beetz, "Towards 3D Point Cloud Based Object Maps for Household Environments," *Robotics and Autonomous Systems,* vol. 56, no. 11, pp. 927-941, November 2008.

[177] Y. Schoenenberger, J. Paratte and P. Vanderg, "Graph-based Denoising for Time-Varying Point Clouds," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, Lisbon, Portugal, July 2015.

[178] H. Su, Z. Duanmu, W. Liu, Q. Liu and Z. Wang, "Perceptual Quality Assessment of 3d Point Clouds," in *International Conference on Image Processing (ICIP)*, Taipei, Taiwan, September 2019.

[179] S. Subramanyam, J. Li, I. Viola and P. Cesar, "Comparing the Quality of Highly Realistic Digital Humans in 3DoF and 6DoF: A Volumetric Video Case Study," in *Conference on Virtual Reality and 3D User Interfaces (VR)*, Atlanta, GA, USA, March 2020.

[180] G. J. Sullivan, J. R. Ohm, W. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 22, no. 12, pp. 1649-1668, September 2012.

[181] J. van der Hooft, M. T. Vega, C. Timmerer, A. C. Begen, F. De Turck and R. Schatz, "Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming," in *International Conference on Quality of Multimedia Experience (QoMEX)*, Athlone, Ireland, May 2020.

[182] G. K. Wallace, "The JPEG Still Picture Compression Standard," *IEEE Transactions on Consumer Electronics,* vol. 38, no. 1, pp. xviii-xxxiv, February 1992.

[183] K. Wolff, C. Kim, H. Zimmer, C. Schroers, M. Botsch, O. Sorkine-Hornung and A. Sorkine-Hornung, "Point Cloud Noise and Outlier Removal for Image-Based 3D Reconstruction," in *IEEE International Conference on 3D Vision (3DV)*, Stanford, CA, USA, October 2016.

[184] X. Wu, Y. Zhang, F. Chun-ling, J. Hou and S. Kwong, "Subjective Quality Study and Database of Compressed Point Clouds with 6DoF Head-mounted Display," *arXiv: Image and Video Processing*, August 2020.

[185] E. Zerman, C. Ozcinar, P. Gao and A. Smolic, "Textured Mesh vs Coloured Point Cloud: A Subjective Study for Volumetric Video Compression," in *International Conference on Quality of Multimedia Experience (QoMEX)*, Athlone, Ireland, May 2020.

[186] J. Zhang, W. Huang, X. Zhu and J.-N. Hwang, "A Subjective Quality Evaluation for 3D Point Cloud Models," in *International Conerance. on Audio, Language and Image Processing*, Shanghai, China, July 2014.

[187] MPEG 3DG, "Draft test conditions and complementary test material," in *ISO/IEC MPEG N16716*, Geneva, Switzerland, January 2017.

[188] 3DGC Group, "Efficient Implementation of the Lifting Scheme in TMC13," in *ISO/IEC JTC1/SC29/WG11 input document m43781*, Ljubljana, Slovenia, July, 2018.

[189] 3DG Group, "G-PCC Codec Description V7," in *ISO/IEC MPEG N19331*, Online, April 2020.

[190] ITU-R Recommendation P.1401, "Methods, Metrics and Procedures for Statistical Evaluation, Qualification and Comparison of Objective Quality Prediction Models," Geneva, Switzerland, July 2012.

[191] "V-PCC Codec Description," in *ISO/IEC JTC 1/SC 29/WG 7 N00012*, Virtual, October 2020.