Quality Evaluation of Point Cloud Coding Solutions

João Pedro Casanova Prazeres

Tese para obtenção do Grau de Doutor em
Engenharia Eletrotécnica e de Computadores
(3º ciclo de estudos)

Orientador: Prof. Doutor António Manuel Gonçalves Pinheiro

Juri:
Prof. Doutor António João Marques Cardoso
Prof. Doutor António Manuel Gonçalves Pinheiro
Prof. Doutor Armando José Formoso Pinho
Prof. Doutor Luís Alberto da Silva Cruz
Prof. Doutora Manuela Areias da da Costa Pereira de Sousa

Provas públicas realizadas a 17 de setembro de 2025.

Declaração de Integridade

Eu, João Pedro Casanova Prazeres, que abaixo assino, estudante com o número de inscrição D2907 de Engenharia Eletrotécnica e de computadores da Faculdade de Engenharia, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o Código de Integridades da Universidade da Beira Interior.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã 6/10/2025

Assinado por: **João Pedro Casanova Prazeres**
Num. de Identificação: 14319898
Data: 2025.10.06 15:39:50+01'00'

Funding

This work was supported by IT internal project PLive X-0017-LX-20 and by FCT - Fundação para a Ciência e Tecnologia, I.P. by project reference UIDB/50008/2020, and DOI identifier https://doi.org/10.54499/UIDB/50008/2020.

List of Publications

Papers included in the thesis resulting from this doctoral research program

1. List of papers accepted for journal publication:

    - **Quality evaluation of point cloud compression techniques**
      Joao Prazeres, Manuela Pereira, Antonio M.G. Pinheiro
      Signal Processing: Image Communication, Volume 128, 2024, 117156, ISSN 0923-5965
      DOI: 10.1016/j.image.2024.117156
    - **Performance Analysis of Deep Learning-Based Lossy Point Cloud Geometry Compression Coding**
      Joao Prazeres, Rafael Rodrigues, Manuela Pereira, Antonio M.G. Pinheiro," in IEEE Access, vol. 13, pp. 76000-76015, 2025,
      DOI: 10.1109/ACCESS.2025.3561895.

2. List of papers accepted for conference publication

    - **Quality analysis of point cloud coding solutions**
      Joao Prazeres, Manuela Pereira, and Antonio M.G. Pinheiro
      Electronic Imaging 34 (2022): 1-6. DOI: 10.2352/EI.2022.34.17.3DIA-225
    - **Subjective and Objective Testing in Support of the JPEG Pleno Point Cloud Compression Activity**
      S. Perry, L. A. da Silva Cruz, Joao Prazeres, Antonio M.G. Pinheiro, Emil Dumic, Davi Lazzarotto, Touradj Ebrahimi
      2022 10th European Workshop on Visual Information Processing (EUVIP), Lisbon, Portugal, 2022, pp. 1-6
      DOI: 10.1109/EUVIP53989.2022.9922803
    - **On the stability of point cloud machine learning based coding**
      Joao Prazeres, Rafael Rodrigues, Manuela Pereira and Antonio M.G. Pinheiro
      2022 10th European Workshop on Visual Information Processing (EUVIP), Lisbon, Portugal, 2022, pp. 1-6
      DOI: 10.1109/EUVIP53989.2022.9922676
    - **Quality Evaluation of Machine Learning-based Point Cloud Coding Solutions**
      Joao Prazeres, Rafael Rodrigues, Manuela Pereira, and Antonio M.G. Pinheiro. 2022.
      Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis (APCCPA '22). Association for Computing Machinery, New York, NY, USA, 57–65.
      DOI: 10.1145/3552457.3555730

- **Subjective Quality Evaluation of Point Clouds with 3D Stereoscopic Visualization**
  Joao Prazeres, Manuela Pereira, and Antonio M.G. Pinheiro
  IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 2022, pp. 2861-2865
  DOI: 10.1109/ICIP46576.2022.9897937
- **JPEG Pleno Call for Proposals Responses Quality Assessment**
  Joao Prazeres, Z. Luo, Antonio M.G. Pinheiro, L. A. da Silva Cruz and S. Perry
  ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5
  DOI: 10.1109/ICASSP49357.2023.10094713
- **JPEG Pleno Learning-Based Point Cloud Coding: A Performance Analysis**
  Joao Prazeres, Rafael Rodrigues, Manuela Pereira and Antonio M.G. Pinheiro,
  2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 1890-1894
  DOI: 10.1109/ICIP49359.2023.10222278
- **Subjective Quality Evaluation Of Point Clouds Using a Head-Mounted Display**
  Joao Prazeres, Rafael Rodrigues, Manuela Pereira and Antonio M.G. Pinheiro
  ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5,
  DOI: 10.1109/ICASSP49660.2025.10889051.

3. List of papers currently awaiting decision

   - **Point Cloud Quality Assessment: Benchmarking Objective Quality Features**
     Joao Prazeres, Rafael Rodrigues, Manuela Pereira and Antonio M.G. Pinheiro
   - **Quality Analysis of the Coding Bitrate Tradeoff Between Geometry And Attributes For Colored Point Clouds**
     Joao Prazeres, Rafael Rodrigues, Manuela Pereira and Antonio M.G. Pinheiro

4. Other publications produced during this doctoral program but not included in the reported research of the thesis:

   - **Quality evaluation of the JPEG Pleno Holography Call for Proposals response**
     Joao Prazeres, Antonin Gilles, Raees Kizhakkumkara Muhamad, Tobias Birnbaum,Peter Schelkens and Antonio M. G. Pinheiro,
     2022 14th International Conference on Quality of Multimedia Experience (QoMEX), Lippstadt, Germany, 2022, pp. 1-6
     DOI: 10.1109/QoMEX55416.2022.9900913

- **Definition of common test conditions for the new JPEG pleno holography standard**
  Antonio M. G. Pinheiro, Joao Prazeres, Antonin Gilles, Tobias Birnbaum, Raees Kizhakkumkara Muhamad, Peter Schelkens
  Proc. SPIE 12138, Optics, Photonics and Digital Technologies for Imaging Applications VII, 121380N (17 May 2022)
  DOI: 10.1117/12.2624499

5. JPEG input documents produced during this doctoral program:

   - wg1m92056-Study on Draco Point Cloud Encoder Performance
   - wg1m93087-Subjective and Objective Quality Evaluation of the Response to the Call for Proposals JPEG Pleno holography
   - wg1m94050-Results of JPEG Pleno Point Cloud Exploration Study 5
   - wg1m95047-Quality Evaluation of Point Cloud Machine Learning based Codecs
   - wg1m96070-Report on the JPEG Pleno CfP subjective quality assessment and objective metrics
   - wg1m97080-JPEG Pleno Point Cloud Report on Metric Performance from CfP
   - wg1m99138-performance D3 metric
   - wg1m99139-subjective evaluation using a HMD equipment
   - wg1m99137-point cloud database analysis
   - wg1m100002-Proposal of a New Point Cloud Test Dataset for JPEG Pleno PCC
   - wg1m100119-FSM Quality Assessment of Point Clouds based on Quality Features Selection
   - wg1m100120-Benchmarking Objective Point Cloud Quality Metrics
   - wg1m101149-Report of Exploration Study 7.1 - Study of Objective Quality Metrics

Agradecimentos

Aos meus pais Nélida Sanches Casanova e Jorge António Cordeiro Prazeres, por todo o apoio.
Ao meu orientador Professor Doutor António Manuel Gonçalves Pinheiro, por toda a paciência e orientação.
À minha namorada Catarina Pereira Bento, por todo o amor e paciência e apoio ao longo destes anos.
Ao meus amigos. especialmente os que se disponibilizaram a deslocar-se ao laboratóio para fazer testes subjetivos.

Resumo

As nuvens de pontos registaram um grande aumento de popularidade. Desde jogos a aplicações médicas, passando por condução autónoma e cartografia urbana, as nuvens de pontos têm sido amplamente utilizadas no mundo tecnológico atual.

À medida que aumenta a procura por conteúdos de nuvens de pontos, aumenta também a necessidade de soluções eficientes de codificação de nuvens de pontos. O acesso a essas soluções é importante para o armazenamento e a transmissão eficientes dos dados das nuvens de pontos, uma vez que estas são normalmente representadas por enormes quantidades de informação.

No entanto, é crucial ter acesso a métodos de qualidade que avaliem com precisão as soluções de codificação de nuvens de pontos. Isto permite aos criadores dessas soluções testar com precisão o seu codificador em vários ambientes diferentes, ajustando o desenvolvimento do codificador conforme os mesmos.

No passado, foram estabelecidos modelos de qualidade subjetivos para avaliar a qualidade de imagens e vídeos. Com base neste conhecimento, foram desenvolvidos novos modelos para a avaliação de nuvens de pontos, embora existam diferenças cruciais devido à natureza tridimensional das nuvens de pontos.

Recentemente, a crescente popularidade dos codificadores baseados em aprendizagem, levou a uma nova análise de desempenho dos modelos de qualidade desenvolvidos, uma vez que as distorções causadas tendem a ser diferentes das criadas pelas tecnologias de codificação tradicionais.

Esta tese tem como objetivo investigar esses modelos de qualidade subjetiva bem estabelecidos, a fim de avaliar o desempenho das soluções de codificação de nuvens de pontos, nomeadamente as que são baseadas na aprendizagem. Além disso, é também importante compreender o desempenho das atuais métricas de qualidade objetiva de nuvens de pontos na avaliação da qualidade das soluções de codificação de nuvens de pontos baseadas na aprendizagem.

Para atingir este objetivo, foram realizados vários estudos de qualidade, sob diferentes condições de visualização, e considerando várias soluções de codificação de nuvens de pontos presentes no estado da arte. Além disso, foi realizada uma extensa análise comparativa de métricas de qualidade objetiva ao longo deste programa de doutoramento, a fim de avaliar o seu desempenho na previsão da qualidade das soluções de codificação de nuvens de pontos baseadas na aprendizagem. Em última análise, isto conduziu a múltiplas contribuições que foram propostas e aceites pela comunidade científica e que foram úteis para compreender o desempenho das soluções de codificação de nuvens de pontos, o impacto da visualização na perceção da qualidade, bem como o desempenho de métricas objetivas de qualidade de nuvens de pontos.

Palavras-chave

Nuvem de pontos, Qualidade Subjetiva, Qualidade Objetiva, Codificação, Medida de Opinião Média, Baseado em Aprendizagem

Abstract

Point clouds experienced a large increase in popularity. From gaming to medical applications, autonomous driving, and urban mapping, point clouds have been widely used in the current technological world.

As the demand for point cloud content increases, the need for efficient point cloud coding solutions also increases. Access to such solutions is important for efficient storage and transmission of point cloud data, because they are typically represented by huge amounts of information.

It is, however, crucial to have access to quality methods that accurately benchmark point cloud coding solutions. This allows the developers of such solutions to accurately test their codec in several different environments, adjusting the codec development accordingly.

In the past, subjective quality models were established to assess the quality of images and videos. Based on this knowledge, new models were developed for point cloud content, though there are crucial differences due to the 3D nature of point clouds. Recently, the growing popularity of learning-based codecs led to a new analysis of the performance of the developed quality models, as the caused distortions tend to be different from those created by the traditional coding technologies.

This thesis aims to research those well established subjective quality models in order to assess the performance of point cloud coding solutions, namely the ones that are learning-based. Furthermore, it was also important to understand how the current point cloud objective quality metrics perform in assessing the quality of learning-based point cloud coding solutions.

To achieve this goal, several quality studies were conducted under different viewing conditions and considering several state-of-the-art point cloud coding solutions. Furthermore, extensive objective quality metrics benchmarking was conducted across this doctoral program in order to assess their performance in predicting the quality of learning-based point cloud coding solutions. Ultimately, this led to multiple contributions that were proposed and accepted by the scientific community and that were helpful in understanding the performance of point cloud coding solutions, the impact of the display on quality perception, and the performance of objective point cloud quality metrics.

Keywords

Point clouds, Subjective Quality, Objective Quality, Coding, Mean Opinion Score, Learning based

Resumo Alargado

A popularidade das nuvens de pontos tem aumentado recentemente. Desde a indústria dos jogos, passando pela condução autónoma até ao mapeamento urbano, as nuvens de pontos são largamente utilizadas atualmente. Com o aumentar da popularidade das nuvens de pontos, aumenta a necessidade de soluções eficientes para as codificar, devido à elevada dimensão usualmente exibida. Essas soluções são importantes para armazenamento e transmissão eficiente das nuvens de pontos.

No entanto, é necessário ter acesso a modelos de qualidade que consigam avaliar precisamente a qualidade de compressão de codificadores com perdas de nuvens de pontos. Com estes modelos, o desenvolvimento de codificadores requer modelos de teste apropriados que permitam uma seleção de tecnologias apropriadas e uma parametrização adequada.

Esta tese tem como objetivo estudar as metodologias a utilizar para avaliar o desempenho de soluções de codificação de nuvens de pontos. No passado, estabeleceram-se modelos de qualidade subjetiva para avaliar a qualidade de imagens e vídeo. Baseado nesse conhecimento desenvolveram-se metodologias de avaliação de qualidade de nuvens de pontos, tendo em consideração a natureza tridimensional deste formato de informação visual. Com a crescente popularidade de codificadores baseados na tecnologia de aprendizagem profunda, esses mesmos modelos necessitam de ser testados e validados para essas mesmas soluções, uma vez que os artefactos gerados por essas soluções diferem bastante dos criados por codificadores tradicionais.

O primeiro capítulo desta tese define o âmbito e o problema onde a tese se enquadra, começando com uma explicação do que são nuvens de pontos. São compostos por coordenadas Cartesianas (x, y, z), e cada coordenada pode conter uma lista de atributos associada, contendo tipicamente uma componente RGB e até um vetor normal perpendiculares à nuvem de pontos nessa coordenada. Outros atributos podem ser a informação de reflectância, informação de um sensor físico, etc. É comum que uma nuvem de pontos contenha milhões de pontos, originando um problema caso se pretenda o seu armazenamento ou transmissão. Adicionalmente, a manipulação de nuvens de pontos requer bastante poder computacional. Consequentemente, aplicações que necessitem deste tipo de tecnologia beneficiam de soluções eficientes para codificar nuvens de pontos. Por sua vez o desenvolvimento e a seleção da tecnologia de codificação requer modelos fiáveis de avaliação subjetiva e objetiva.

Há dois métodos principais para avaliar a qualidade de soluções de codificação de nuvem de pontos, nomeadamente avaliação subjetiva e avaliação objetiva. A avaliação subjetiva é usualmente efetuada num ambiente controlado, com as condições definidas na norma ITU-R BT.500-15.

A avaliação de qualidade de nuvens de pontos difere bastante da avaliação de imagens ou

vídeo 2D, devido à sua inerente representação tridimensional. Uma imagem 2D é composta por pixeis mapeados numa grelha, sem qualquer tipo de informação de profundidade, e sem espaços vazios entre os mesmos. No caso das nuvens de pontos, a informação geométrica é mapeada no espaço 3D. Todavia esta informação não é densa, havendo espaços vazios entre as coordenadas 3D. A natureza esparsa das nuvens de pontos leva a dificuldades acrescidas na sua avaliação de qualidade. Na avaliação subjetiva podem surgir situações indesejáveis, em que o interior dos objetos representados é visualizado devido ao espaçamento entre os pontos, causando uma má perceção de qualidade. Contrariamente às imagens 2D, que contém apenas uma vista, como as nuvens de pontos representam informação 3D, podem ser visualizadas a partir de diferentes ângulos e distâncias. Este facto representa um desafio enorme para a avaliação de qualidade subjetiva, já que em princípio tem de se garantir que a nuvem de pontos seja totalmente visualizada. Caso contrário, pode acontecer que alguns artefactos poderão ficar ocultos durante a avaliação de qualidade subjetiva.

Para se levar a cabo um estudo de qualidade subjetiva, é normalmente considerado que pelo menos 15 sujeitos avaliem cada conteúdo. Depois, a média total das avaliações é calculada para cada estímulo, obtendo as *Mean Opinion Scores*. Esses valores são mapeados com a *bitrate*, que é usualmente quantificada em bits por ponto.

Planear e executar um estudo de qualidade subjetiva requer um planeamento cuidadoso, e exige disponibilidade temporal, e por vezes pode arcar custos financeiros. Como alternativa, durante o desenvolvimento de soluções de codificação é habitual recorrer-se a métricas de qualidade objetiva que de alguma forma medem a fidelidade do sinal ou simulam as avaliações subjetivas. Desta forma, evitam-se os custos temporais relacionados com a avaliação subjetiva. Todavia, para perceber até que ponto estas métricas representam os resultados subjetivos e a perceção humana de qualidade, é necessário avaliar o desempenho das métricas, usando como base resultados da avaliação subjetiva.

Para avaliar as métricas objetivas, usa-se a norma ITU-T P.1401 que estabelece uma metodologia de comparação da avaliação subjetiva com os resultados da métricas, usando como indicadores estatísticos a correlação de *Pearson* e *Spearman*, o *root-mean-squared error* e o *ratio* de *outliers*.

Esta tese beneficiou ainda do envolvimento no processo de normalização *JPEG Pleno Learning Based Point Cloud Coding*, que visa a criação de uma norma para a codificação de nuvens de pontos, utilizando tecnologia de aprendizagem profunda.

Este programa de doutoramento focou-se maioritariamente em três áreas diferentes, nomeadamente a avaliação de qualidade subjetiva, a avaliação de qualidade objetiva, e a avaliação do desempenho de de codificadores baseados em aprendizagem profunda.

No âmbito da avaliação de qualidade subjetiva, os trabalhos desenvolvidos comparam diferentes soluções de codificação. Vários estudos foram efetuados para avaliar o desempenho de vários codificadores de nuvens de pontos, nomeadamente utilizando um ecrã

2D, um ecrã 3D estereoscópico, e um *head-mounted display* (HMD). Os estudos não apresentaram diferenças estatisticamente relevantes, permitindo a validação para ambientes imersivos do protocolo tipicamente utilizado. Foi ainda verificado que efetuar estudos de qualidade subjetiva em ambientes imersivos apresenta vantagens significativas.

Adicionalmente, um estudo preliminar permitiu preparar a *Call for Proposals* estabelecida durante o processo de normalização. O objetivo foi estudar o desempenho dos codificadores âncora selecionados (neste caso, as normas de MPEG, V-PCC e G-PCC), assim como as métricas de qualidade objetivas selecionadas pelo comité da JPEG.

Os estudos de qualidade subjetiva são tipicamente complementados com uma avaliação objetiva. Várias métricas objetivas foram avaliadas, e um estudo estatístico para analisar as diferenças entre o desempenho das mesmas foi efetuado. Adicionalmente, também foi estudada a influência do método de cálculo dos vetores perpendiculares à nuvem de pontos em cada ponto que são utilizados em algumas métricas objetivas. Esse estudo foi ainda complementado com uma análise da parametrização destes modelos.

Adicionalmente, foi efetuado um estudo da efetividade e importância dos descritores, definidos em várias métricas para a definição da qualidade subjetiva. Estes descritores foram analisados, e a importância dos mesmos para a estimação final de qualidade foi ordenada utilizando o *Recursive Feature Extraction* (RFE). O estudo levou à definição de um modelo baseado na combinação de descritores para a avaliação de qualidade *Feature Selection Metric* (FSM), que foi comparado com as métricas de estado da arte em quatro bases de dados diferentes.

No âmbito da análise de soluções de aprendizagem profunda, foram efetuados vários estudos que permitiram entender o desempenho das soluções de aprendizagem profunda. Inicialmente foi efetuado um estudo que avaliou apenas soluções que codificam geometria, comparadas com o codificador G-PCC normalizado pela MPEG, e uma solução que representa uma evolução do mesmo. O estudo foi complementado com a análise de várias métricas objetivas.

Adicionalmente, um estudo de qualidade subjetiva foi efetuado no âmbito da *JPEG Pleno Learning Based Point cloud Coding Call for Proposals* para avaliar os codificadores submetidos. Este estudo avaliou codificadores de aprendizagem profunda que codificam a informação de cor e geometria. Os resultados proporcionados pelos codificadores propostos foram comparados com os definidos nas normas da MPEG, o G-PCC e o V-PCC. O codificador com o melhor desempenho foi selecionado para ser a base de desenvolvimento da norma *JPEG Pleno Learning based Point Cloud Coding*.

Adicionalmente, foram efetuados estudos sobre o desempenho dos codificadores de aprendizagem profunda durante o treino e a sua estabilidade. Os estudos reportam a evolução de qualidade durante o processo de treino de várias soluções de aprendizagem profunda. Para o estudo que analisa codificadores que apenas codificam geometria, foi utilizada uma métrica que apenas considera a informação de geometria. Para o segundo estudo, que

considera um codificador para a geometria e cor da nuvem de pontos simultaneamente introduziram-se métricas que também consideram a cor.

O segundo capítulo da tese apresenta uma revisão de literatura nas três principais áreas em que a tese se enquadra, nomeadamente a avaliação de qualidade subjetiva, avaliação de qualidade objetiva e codificação de nuvens de pontos. O capítulo apresenta ainda metodologias de aquisição das mesmas, incluindo métodos como a fotogrametria, captação com sensores LiDAR, etc.

No âmbito da avaliação de qualidade subjetiva, é inicialmente apresentado o surgimento da avaliação de qualidade subjetiva, e os acontecimentos que levaram ao desenvolvimento dos primeiros protocolos de avaliação. É depois feita uma introdução ao campo específico de avaliação de qualidade subjetiva de nuvens de pontos, onde é inicialmente feita uma revisão de trabalhos relevantes na área.

Seguidamente é apresentada uma revisão da literatura dos métodos de avaliação de qualidade objetiva. Estes métodos dividem-se em três grandes grupos:

- Métricas com referência completa
- Métricas com referência reduzida
- Métricas sem referência

As métricas com referência completa comparam a totalidade das nuvens de pontos original e a distorcida. Estas métricas ainda podem ser divididas em dois tipos: As que medem a fidelidade do sinal, e as que tentam simular o sistema humano de visualização. As métricas com referência reduzida extraem apenas um pequeno conjunto de descritores, de ambas as nuvens de pontos e estimam a qualidade exclusivamente na comparação destes descritores. As métricas sem referência consideram apenas a nuvem de pontos distorcida, e baseiam-se no facto de que um humano consegue avaliar a qualidade de um conteúdo distorcido sem ver ao original.

As métricas de avaliação objetiva podem ainda considerar apenas a informação de geometria, a geometria e o vetor normal (representa a direção perpendicular à nuvem de pontos numa coordenada específica), apenas a cor, ou considerar a informação de geometria e de cor. A tese ainda reporta os modelos de avaliação do desempenho das métricas de qualidade objetiva, com uma explicação da função de *fitting*, assim como os indicadores estatísticos para avaliação de desempenho.

A secção seguinte apresenta o estado da arte relativo à codificação de nuvens de pontos. Os codificadores podem ser divididos em três grandes grupos, os codificadores baseados na segmentação de geometria (sendo os baseados em *octree* os mais populares), baseados em projeções, e os baseados em tecnologia de aprendizagem profunda.

Os codificador definido pela MPEG denominado de *Geometry-Based Point Cloud Compression* (G-PCC), foi desenvolvido para nuvens de pontos estáticas. O codificador baseia-

se na representação *octree*, permitindo ainda a utilização do método *trisoup* na codificação da geometria. Os atributos podem ser codificados com três metodologias diferentes, nomeadamente a *Region Adaptive Hierarquical Transform* (RAHT), *predicting transform* e *lifting transform*. Esta última é baseada na *prediting transform*, sendo a junção das duas referida como *predicting/lifting transform*.

Seguidamente explica-se o codificador *Video Point Cloud Coding* (V-PCC). O codificador gera projeções 2D das nuvens de pontos, que são depois codificadas utilizando codificadores de vídeo.

É dedicada também uma especial atenção nos codificadores baseados na tecnologia de aprendizagem profunda. Apresentam-se seguidamente os codificadores que apenas codificam geometria, uma pequena revisão dos que codificam apenas a cor, e uma explicação detalhada do standard da *JPEG Pleno Learning Based Point Cloud Coding*, que é o único *codec* que codifica a cor e a geometria.

O capítulo seguinte apresenta as metodologias que foram utilizadas durante o desenvolvimento da tese, nomeadamente as considerações a ter quando se escolhe um conjunto de nuvens de pontos para levar a cabo a avaliação subjetiva. o capítulo apresenta ainda exemplos de nuvens de pontos codificadas com codificadores utilizados durante a tese, e apresenta também as recomendações de *bitrate* alvo a utilizar neste tipo de avaliações.

Seguidamente, descrevem-se os métodos de avaliação subjetiva utilizados. Mais detalhadamente, apresenta-se a metodologia para gerar vídeos em 2D, 3D e detalhes de visualização utilizando o *software Unity*. Refere-se também o método de manipular a nuvem de pontos para que a parte oposta da nuvem não fique visível. Finalmente apresentam-se os métodos de computação do vetor normal, e os vários testes estatísticos para avaliar as diferenças estatisticamente relevantes entre avaliações subjetivas e métricas de avaliação objetivas. Seguidamente, apresentam-se as conclusões onde se destacam os comentários sobre avaliação de qualidade subjetiva e objetiva. São discutidas também as soluções de codificação baseadas em aprendizagem. Finaliza-se com uma análise do trabalho futuro.

O último capítulo da tese apresenta os artigos efetuados durante o programa doutoral, aceites para publicação.

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

ADLPCC	Adaptive Deep Learning Point Cloud Coding
AE	Autoencoder
AR	Augmented Reality
AUC	Area Under ROC Curve
BD	Bjontegaard Delta
bpp	bits per point
BRDF	Bidirectional Reflectance Distribution Function
CCp	Correct Classification Percentage
CfP	Call for Proposals
CNN	Convolutional Neural Network
CTC	Common Test Conditions
CTTC	Common Training and Test Conditions
COL	Color
DL	Deep-Learning
DSCQS	Double Stimulus Continuous Quality Scale
FL	Focal Loss
FR	Full-Reference
FSM	Feature Selection Model
GeoCNN	Improved Deep Point Cloud Geometry Compression
GEO	Geometry
GEO + COL	Geometry plus color
GEO + LUM	Geometry plus luminance
GPCC	Geometry Point Cloud Compression
HMD	Head Mounted Display
HSV	Human Visualization System
IRB	Inception-Residual Blocks
ISO	International Organization for Standardization
IEC	International Electrotecnical Commission
IT	Instituto de Telecomunicações
ITU-T	International Telecommunication Union Telecommunication
JPEG	Joint Photographic Experts Group
KD-tree	k-dimensional tree
KNN	K Nearest Neighbors
LD	Local Density
LiDAR	Light Detection and Ranging
LUM	Luminance
LUT_SR	Look Up Tables Super Resolotion
MOS	Mean Opinion Scores

MPEG	Moving Picture Experts Group
NR	No-Reference
OR	Outlier Ratio
PC	Point Cloud
PCC	Pearson Correlation Coefficient
PCGC	Multiscale Point Cloud Compression
PSNR	Peak Signal Noise Ratio
RBF	Radial Based function
ReLU	Rectified Linear Unit
RFE	Recursive Feature Elimination
RFG	Random Forest Generator
RMSE	Root Mean Squared Error
ROC	Receiving Operating Characteristic
RiR	Ridge Regression
RR	Reduced-Reference
RSDLPCC	Resolution Scalable Deep Learning Point Cloud Coding
SfM	Structure from Motion
SGD	Stochastic Gradient Descent
SROCC	Spearman Rank Order Correlation Coefficient
SSCQE	Single Stimulus Continuous Quality Evaluation
SSIM	Structural Similarity Index
SR	Super-Resolution
SVR	Support Vector Regression
ToF	Time-of-flight
UBI	Universidade da Beira Interior
VAE	Variational Autoencoder
VM	Verification Model
VPCC	Video Point Cloud Compression
VR	Virtual Reality

Chapter 1

Introduction

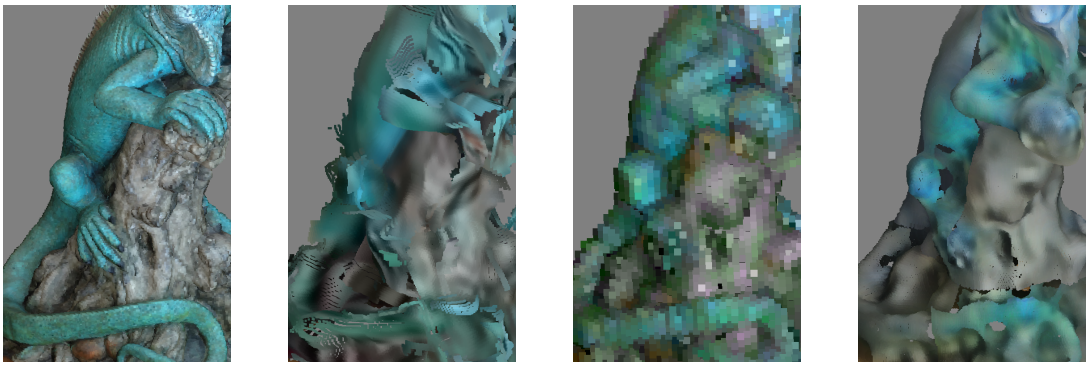
1.1 Overview

This doctoral study aims to study, analyze, and develop methodologies for the quality evaluation of point cloud lossy coding solutions. Lossy coding refers to codecs that do not maintain the original visual information, and somehow change it to reduce the required amount of memory for its representation. However, this representation should keep a perceptual representation as close as possible to the original.

Subjective quality evaluation is the most reliable methodology to assess the perceptual differences between the coded representation and the original visual information. It is conducted by coding a representative selection of original content with the coding solutions under test. Usually, coding solutions are benchmarked against a set of anchor codecs that are standardized or recognized as state of the art (such as MPEG V-PCC and G-PCC in the case of point clouds). The coded content usually represents different levels of quality, usually from very low to very high. In some cases, mathematically or visually lossless qualities are also analyzed.

During the development of a codec, it is impossible to conduct subjective quality for each development stage, as it is very time-consuming, demands very careful planning, and in some cases becomes quite expensive. Therefore the quality evaluation process uses objective quality metrics. These metrics might represent a measure of the signal fidelity or, as is desirable, somehow try to mimic the human perception of quality. Objective quality evaluation metrics are usually validated based on subjective evaluation.

The two main traditional point cloud coding methodologies are either based on projections or on octrees [1]. Recently, the increasing popularity of learning-based technology led to the development of several point cloud coding solutions based on their usage [8, 7, 6]. As observed in Fig. 1.1, learning-based coding solutions result in different types of artifacts, depending on the coding technology. Therefore, measuring the performance of point cloud coding solutions becomes challenging, as quality models need to adapt to the different types of distortions. The initial quality evaluation methodologies were developed for the evaluation of octree or projection-based solutions. Since the artifacts generated by them are different from the ones created by learning-based solutions, it is necessary to validate or to adapt those methods to the new developments or, eventually, to develop entirely new quality evaluation models. Finally, point clouds are widely used in immersive environments (AR/VR) and protocols, and the existing protocols and



(a) Reference (b) Projection-based (c) Octree-based (d) Learning-based

(e) Projection-based (f) Octree-based (g) Learning-based

Figure 1.1: Examples of the different artifacts generated by point cloud coding solutions. The upper row depicts examples of low-quality reconstruction, while the lower row shows examples of high-quality reconstruction.

methods should be evaluated for such conditions.

1.2 Thesis focus and scope

Point clouds became one of the most popular methods for representing volumetric data [22]. They consist of a set of Cartesian coordinates (x, y, z) that may have an associated list of attributes, such as an RGB component, reflective information, physical sensor information, or normal vectors [23]. This leads to point cloud data containing an enormous amount of data [24], as they might contain millions of points. This presents a problem if there is a need to store or transmit a point cloud over a network. Furthermore, handling large point clouds requires significant computational effort. As a result, applications that use this popular representation model benefit from effective point cloud encoding methods, as well as precise subjective and objective quality models for benchmarking.

1.2.1 Point Cloud Use Cases

Point clouds have a wide number of applications, from gaming and medical applications to cultural heritage preservation and autonomous vehicles. A number of use cases are identified in the JPEG Pleno Point cloud use cases and requirements document [2], and the most relevant are described here.

1.2.1.1 Virtual, augmented and mixed reality

Point clouds play an important role in supporting the display of 3D content in virtual, augmented, and mixed reality environments [2]. That content can be generated using CAD or 3D scanning equipment. The resolution and number of points will vary, depending mainly on the type of object or 3D scanner. Point clouds that were created using CAD software will most likely be arranged on regular grids and patterns, in contrast to data collected with 3D scanning, which will be arranged in an irregular geometric pattern. Important data attributes used in these applications are color, gloss, or texture maps, as well as a bidirectional reflectance distribution function (BRDF).

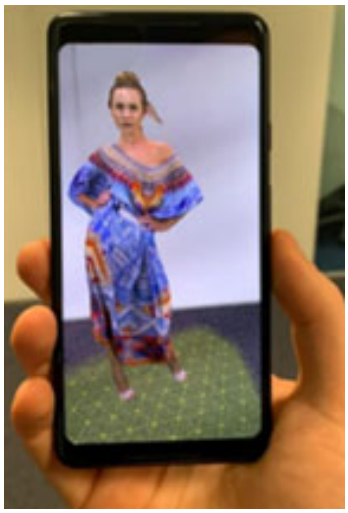


Figure 1.2: Example of AR for point clouds (Figure taken from [1])

1.2.1.2 Design, manufacturing and 3D printing

This type of use case mainly involves point clouds specifically created to support the content for 3D printing and traditional manufacturing (Fig. 1.3). The data will most likely be arranged on regular grid patterns associated with CAD software; the resolution of the point will be highly dependent on the type of object and the industry in which the object will be used. The point cloud attributes should include color and material appearance information, most likely linked to databases containing information about used materials. Since this case requires accurate representation of point clouds, lossless representation

will be important. Additionally, support for attributes with lossless coding will likely be needed in the cases where the attribute looks to an external database. To protect intellectual property, data privacy and security should be linked to the content, with some form of encryption.

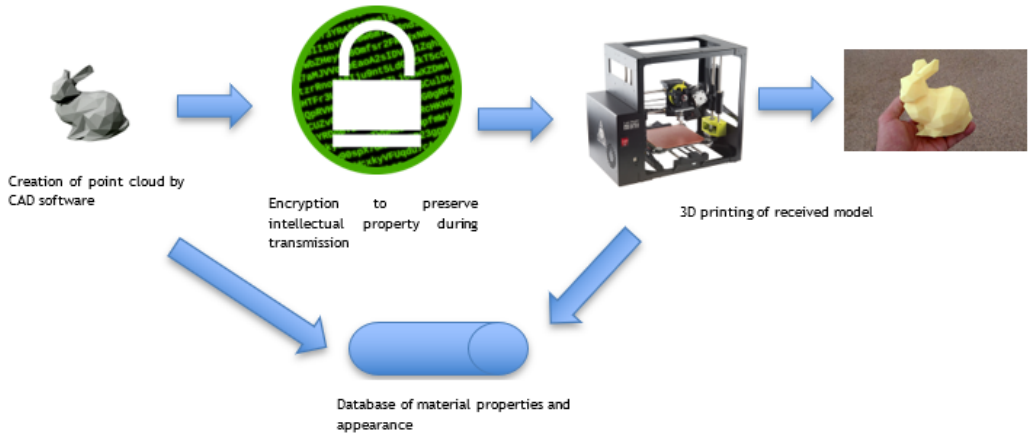


Figure 1.3: Usage of CAD created point cloud for 3D printing (Figure taken from [2])

1.2.1.3 3D medical image

Refers to the point clouds obtained by 3D scanning of internal or external human or animal anatomy with the purpose of research, medical record keeping, diagnosis, and preparation of treatment. This type of point cloud needs to be encoded in a way that preserves the ability for subsequent analysis, thus requiring an accurate representation of point clouds, meaning that lossless or very low coding will be prioritized. It also may require multiple scans of anatomical structures and internal detail, potentially needing to keep point clouds from individual scans distinct in the coded format. To support examination by medical professionals, view selectivity, region of interest, and resolution granularity/scalable bit streams capabilities will most likely be important. Although some point attributes may be encoded by a lossy method, support for attributes with lossless coding will be needed where such attributes are crucial for diagnosis and analysis. There should also be a requirement to provide means to guarantee the privacy and security needs associated with content.

1.2.1.4 Prosthesis and body parts design

This use case targets point clouds that were created to support the production of 3D prosthetics. In this use case, the point cloud resolution, as well as the number of points, is dependent on the type of object. As the point clouds will mainly be generated by CAD software, they are likely to be arranged in regular grids. However, if the point cloud is generated by scanning a given body part, it may be arranged in an irregular pattern, as

well as being quite sparse. Furthermore, the generated point cloud will also have associated attributes linked to the scanned body part. These attributes are usually color, as well as material appearance, most likely linked to a database that contains a set number of materials.

1.2.1.5 3D scanning for project management support

This use case targets point clouds resulting from the scanning of 3D construction materials in order to support the visualization of architecture and infrastructure of a building site or the surrounding area. The content can be generated by CAD design or generated through scanning of the physical object. Additionally, there may be cases where the generated point cloud is a merge of both methods. The resolution and number of points will be dependent on the type of industry and object. As in the other use cases, the point clouds that are generated by CAD will be organized in regular shapes, while the ones resulting from scanning equipment will be more irregular. The associated attributes are likely to be color information, gloss, bump maps or texture maps, as well as material appearance information.

1.2.1.6 Consumer Retail

This use case concerns point clouds that are used to store the 3D shape and details of objects. This use case can be divided in three types of objects:

- Small objects, such as jewelry, decoration, shoes, etc.
- Medium sized objects, such as as cars, motorcycles or furniture.
- Large objects, such as houses or apartments.

The point clouds that match this use case are often complicated in structure, as well as a complex material appearance. The resolution will be very dependent on the type of object, as some point clouds may a mixture of individually scanned objects that are merged into a single point cloud entity.

1.2.1.7 Cultural Heritage

This use case targets the preservation of objects that in some way depict cultural heritage, such as pottery, bones, paintings, statues, etc.

The point clouds that depict small artifacts (such as pottery) and medium-sized artifacts (such as statues) are usually complex in structure, with a high geometric resolution and a complex material appearance. These point clouds often serve the purpose of enabling a

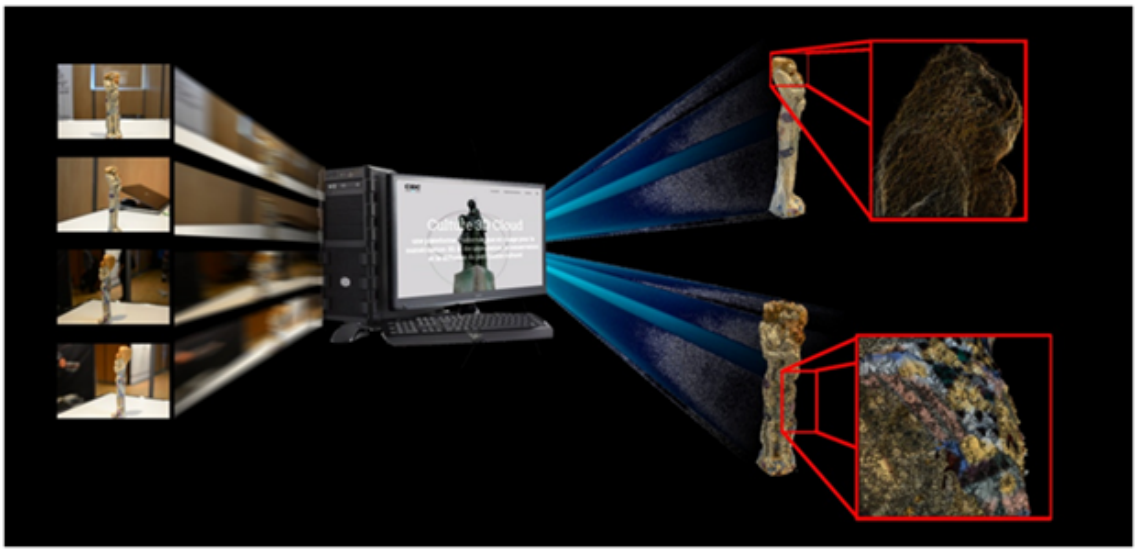


Figure 1.4: Cultural heritage in the 3D Cloud project (http://c3dc.fr/) (Figure taken from [3])

detailed analysis and being visualized by the public on either websites or mobile devices. Point clouds that depict paintings or murals have three main goals. Recognize the artist that created the work, understand how it was created, and most importantly, detect forgeries. Multiple scannings of the object may be required, and they usually contain complex texture maps. There is also the possibility that the point clouds are represented as a height map above the surface in order to reflect the flat geometry of paintings and murals.

Point clouds that depict facades of monuments also require multiple scans in order to capture the entire object. The generated point clouds are likely to contain very complex structure and material appearance.

The scannings will most likely be conducted under conditions of uncontrolled lighting.

1.2.1.8 Remote sensing and geographical information systems

This use case targets the recording of 3D shapes and details of landscapes, either urban or rural. These point clouds will be acquired under uncontrolled lighting, and multiple scans may be required to fully reconstruct the scene. In this particular use case, it is not always necessary to capture the color information. Fig. 1.5 shows a point cloud that falls into this use case. It can be observed that the data has high resolution but does not contain information on the material appearance of the scene.

1.2.1.9 Autonomous vehicles, drones

The main purpose here is to spot objects such as obstructions or pedestrians. Resolution may vary with the direction in which the vehicle is moving, and it may or may not require

Figure 1.5: An example of a point cloud of wide area scanning. The data is of high resolution, but without accurate material appearance information or color. The color in this figure represents height and is not related to the color of the imaged scene. The image is a LIDAR scan of Buckingham Palace, UK and is courtesy of Environmental Agency (https://www.flickr.com/photos/environment-agency/27489358013) [CC BY 2.0 (https://creativecommons.org/licenses/by/2.0/)] (Figure taken from [1]).

accurate color and material appearance. The point clouds require lossless or very high-quality lossy coding. The representation should preserve sharp edges and fine details, and requirements for real time and low latency are necessary, as there is a need to collect and analyze incoming data continuously.

Figure 1.6: Usage of point cloud in autonomous vehicles (Figure taken from [4])

1.2.1.10 Surveillance

The goal of this use case is to spot objects of interest, such as intruders that are committing some sort of trespassing or trapped survivors in an emergency situation. In this use case, the point clouds are acquired in extremely irregular environments, although point clouds in this use case may not require color information.

1.3 Problem Statement and Goals

Evaluating the quality of 3D data information is inherently different from evaluating 2D images. As shown in Fig. 1.7, a traditional 2D image consists of pixels mapped in a grid without any kind of depth information. In volumetric images such as point clouds, the geometry information is spread through a three-dimensional space, being represented by points that do not represent continuous information. Typically, for visualization, these points are replaced by spheres with a given diameter. Point clouds represented by the spheres are likely to create empty spaces in its volumetric representation. This presents a challenge in assessing the quality of point cloud coding solutions. If the subject assessing the quality of the point cloud content can visualize the interior of the point cloud, it will most likely result in a poor-quality perception.

Contrary to 2D images, which contain only 1 view, point cloud images can be visualized from any point. This presents another challenge, as the point cloud needs to be fully visualized. If this is not done, some artifacts may remain hidden during subjective evaluation.

(a) 00002 (JPEG AIC dataset)

(b) Longdress (JPEG Pleno Database: https://plenodb.jpeg.org)

Figure 1.7: Example of the differences between a 2D image (1.7a) and a projection of a point cloud (1.7b).

As referred to previously, there are two main possibilities to evaluate quality, notably subjective and objective quality evaluation. Subjective quality evaluation is considered the most reliable methodology for quality evaluation, as it provides the most accurate information on the impact of the artifacts generated by the coding solutions. However as previously stated, subjective quality evaluations require very careful planning, are very time-consuming, and eventually become quite expensive. Typical evaluation sessions are usually conducted in a controlled environment, and the conditions are defined in ITU-R BT.500-15 [25].

For a reliable subjective quality evaluation, it is considered that at least 15 subjects must rate each individual stimulus. Then, the average of the scores for each stimulus is com-

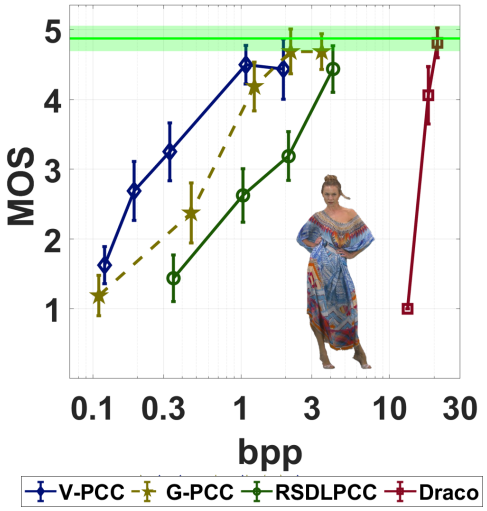


Figure 1.8: Example of an MOS vs bpp plots obtained for the *Longdress* point cloud. The y axis represents the MOS, while the x axis represents the bpp - bits per point (Plot and figure taken from [5]).

puted, resulting in a collection of mean opinion scores (MOS). For a codec evaluation, the MOS values are usually plotted against the bitrate (usually expressed in bits per point (bpp)).

An example can be found in Fig. 1.8. The y-axis represents the obtained MOS for the *Longdress* point cloud. The x-axis represents the bitrate in bits per point (bpp). Each line represents a codec. The vertical continuous lines represent the 95% confidence interval, assuming a Gaussian distribution. The green horizontal line represents the MOS obtained with the hidden reference pair (comparison of the reference point cloud with itself, without the knowledge of the subjects), and the green shade around it represents its 95% confidence interval.

As it is impossible to conduct subjective quality evaluations during each stage of a codec development, this process must rely on objective quality measures. These metrics should allow a reliable quality evaluation of the distortions introduced by the coding process.

To evaluate the reliability and effectiveness of these metrics, they need to be benchmarked to assess how accurately they represent the results of the subjective quality evaluations. This process uses subjective evaluation scores, considering the methodology and statistical indicators recommended by ITU-T P.1401 [26].

Across this thesis, several methodologies for quality evaluation of point cloud coding solutions will be considered and discussed, namely:

- Conducting subjective quality studies in immersive environments;
- Comparing the performance of several state-of-the-art objective point cloud quality metrics;

- Studying the influence of each metric features in the performance of the objective evaluation and understand which features are the most prominent ones.

Finally, the performance of several point cloud coding solutions is also discussed. To achieve this goal, a list of objectives was defined.

- Study the most widely used methods for subjective quality evaluation of point clouds, and implement them in order to analyze the identified solutions.
- Establish a method to evaluate point cloud coding solutions that only encode geometry information only.
- Develop quality methods that are able to accurately evaluate the quality of point cloud coding solutions in immersive environments.
- Analyze several point cloud properties, in order to construct diverse datasets appropriate for quality evaluation. Such properties include, sparsity, homogeneity, density, color gamut volume and geometry precision.
- Benchmark objective quality metrics, in order to understand the ones that provide the most accurate representation of the subjective evaluation results.
- Analyze the representation provided by the features defined in the best performing point cloud quality metrics.
- Identify the state-of-the-art point cloud coding solutions, including the most prominent types of solutions, octree based coding, projections based coding and learning based coding.

1.3.1 Involvement in Standardization

The research development reported in this thesis benefited from the involvement of the research group on the standardization project JPEG Pleno, notably the development of a static Point Cloud coding standard.

During the doctoral program, several inputs on point cloud quality evaluation models were produced for the JPEG Committee. They were crucial for the quality evaluation process required for the development of the JPEG Pleno Learning based Point Cloud coding standard.

The standardization process typical of the JPEG Committee consists of multiple tasks that must be executed in order to establish a final standard.

Initially, a use cases and requirements document is established [2]. This document identifies the main applications and defines a set of use cases.

The document also specifies the characteristics of point clouds for each use case. Then it establishes a set of requirements for a codec to accomplish the identified use cases.

Afterwards, a common test conditions (CTC) document [27] is created. This document contains information on the evaluation of the coding proposals, notably the subjective quality evaluation protocol and the objective quality metrics that will be employed to assess the performance of the proposed solutions.

Usually, the document also provides information on the data selected for testing. However, in cases where learning based codecs can be proposed, the testing data needs to be hidden, and proposals should not have that information available.

Furthermore, the CTC document also contains information on the anchors and their computation. Those anchors and the proposed solutions are tested between them to provide an analysis of how effective the proposals are considering the state of the art.

Although optional, in some cases a Call for Evidence is then issued [28]. This stage allows the identification and the assessment of possible future proposals.

In the case of the JPEG Pleno Point Cloud coding, there was a single response proposing a learning based codec.

Then it was decided that the call for proposals should be focused on learning-based codecs. Furthermore, to ensure that all the codecs are tested equally, a Common Training and Test Conditions (CTTC) [17] document was defined.

It complemented the initial CTC with a training dataset that all proponents should use. In addition, no information on the testing data was given.

A Call for Proposals (CfP) [29] is then issued, which contains information related to the scope of the standard, and is disseminated in order to encourage researchers and industry to propose their solutions to the JPEG Pleno committee.

The responses to the call for proposals are benchmarked against the previously mentioned anchors and between them [10]. The benchmarking process of the responses to the call for proposals comprises both subjective and objective evaluation and also a validation of the objective metrics. The solution with the best performance and that better respects the testing conditions defined in the CTC document (or the CTTC document in the case of the call for Point Clouds coding of JPEG Pleno) is selected as the basis for the future standard.

Usually the selected proposal is used for the establishment of a Verification Model (VM) [30], which becomes the testing base for the standard development. This VM will be enhanced with other identified technologies in order to improve the performance of the codec. For that, core experiments and exploration studies are defined. Core experiments evaluate if a certain technology should be considered for the standardized solution. Exploration studies aim to verify new methodologies or technologies that might be considered for the standardized solution.

1.4 Main Contributions

The developed research targeted the quality evaluation of point clouds encoded by the different coding solutions. Both subjective and objective models were considered. The developed research also considered the recent development of the new learning based codecs, that result in different distortions and require an adaptation of the quality models.

1.4.1 Subjective quality evaluation

This thesis contributes with a quality comparison of different coding solutions using subjective quality evaluation. Initially, the established coding methodologies based on octree or on point cloud projection compression were considered. Then, as a result of the trend in the development of learning based codecs, several studies on the performance analysis of these types of technologies were performed. The quality studies considered different types of visualization.

Several experiments were conducted to evaluate the performance of point cloud coding solutions under different viewing conditions, notably using a 2D display [16], a 3D stereoscopic display [14], as well as a head-mounted display [31]. It was revealed that there were no statistical differences between the visualization modalities. Furthermore, it was concluded that conducting subjective evaluation using head mounted displays has increased benefits.

During the research development, several subjective evaluations were made to support the JPEG Pleno Point Cloud Coding activity in collaboration with the JPEG Committee. This work was essential for the development of the standardized solution, as it considered the definition of the evaluation protocol for the evaluation of the submitted technologies, the settings and performance of the anchors, and finally the evaluation of the responses to the call for proposals for the selection of the proposed technologies.

1.4.2 Objective Quality Evaluation

Objective quality metrics aim to assess the quality of point cloud coding solutions, by using a mathematical model that predicts quality scores. Those metrics need to be validated using the scores obtained with subjective quality evaluation.

Across the doctoral program, several objective point cloud quality metrics were benchmarked, and their performance reported. The studies conducted [16, 14, 13, 10, 32] considered a set of widely used point cloud quality metrics that either considered, geometry information, color attributes or both. It was observed that the metrics that consider both geometry information and color attributes generally achieve a superior representation of the subjective results, revealing the importance of both geometry and point attributes (in

the considered cases RGB components associated to each point) in the definition of the perceptual quality.

A study conducted during the program [5] considered a wide range of state-of-the-art point cloud quality metrics. Notably, the study included full-reference, reduced-reference, and no-reference metrics, and the respective parametrization of the metrics as well. The models used for the point cloud normal vector information computation for the metrics that rely on them were also analyzed. To further assess the performance of the metrics, a statistical study was also conducted.

The conducted studies provided insight on the performance of several objective metrics, and help to understand the most reliable metrics that should be considered during the development of a codec. The information learned with these studies strongly influenced the development of the JPEG Pleno Point Cloud Learning based codec as it was instrumental for the evaluation of the technologies tested in the different core experiments and explorations studies.

1.4.3 Performance Analysis of Learning-based Coding Solutions

Several studies were conducted to provide a better understanding of the performance and stability of these learning-based solutions.

Initially, a study was conducted to assess three state-of-the-art learning-based solutions that only coded geometry [13]. They were compared to the MPEG standard G-PCC [33], as well as an evolution of it [34].

A subjective quality study was conducted, complemented with the benchmarking of several objective quality metrics. This study helped to understand how the commonly used objective quality evaluation metrics behave in the presence of learning-based artifacts.

A subjective quality evaluation assessing the solutions submitted to the JPEG Pleno Learning Based Point Cloud Coding call for proposals was also conducted [10]. The study benchmarked three learning-based codecs that encode both geometry information and color attributes. They were compared to the existing point cloud coding standards, namely V-PCC and G-PCC [35]. The best-performing solution was selected to be the base of the JPEG Pleno Learning-Based Point Cloud Coding Verification Model.

The performance of learning-based codecs during the training stage was also studied [32]. The conducted research reported the evolution of coding quality during the training process of two different learning-based point cloud compression solutions. As the solutions only encode geometry information, the widely used PSNR MSE D1 [19] metric was employed.

The goal was to assess the differences between the results produced by three different training sessions in the performance of the codecs.

A similar study was conducted for the JPEG Pleno Learning Based Point Cloud Coding Verification Model [36]. The study followed the same procedure, but it also included metrics that considered the color information, namely PCQM [11] and GraphSIM [37]. This analysis was necessary because the codec also encodes the color information. Furthermore, the final working point of each training session was compared to the default codec by computing the Bjontegaard Deltas [38] for each session, using the default codec as a reference.

1.5 Thesis Organization

This thesis is divided into four chapters.

Chapter 1 aims to introduce the thesis goals and briefly present the work conducted during the doctoral program.

Chapter 2 introduces concepts of quality evaluation, namely presenting a review of several works conducted in order to conduct subjective quality evaluation studies, as well as the several objective point cloud quality metrics available in the literature.

Chapter 3 presents the several methodologies for subjective and objective quality evaluation. The chapter also explains the different methods to evaluate statistical differences between subjective quality evaluation studies and metrics.

Chapter 4 presents the conclusions that can be withdrawn, as well as future work.

Chapter 5 consists of the papers published during the doctoral program.

Chapter 2

State of The Art

In this chapter, the state of the art on point cloud technology is described. The following subjects will be reviewed:

- Usual methods for the acquisition of point cloud images.
- Existing coding solutions, notably the existing standards, as well as learning-based coding solutions.
- Revision of subjective quality evaluation protocols.
- Revision of objective point cloud quality metrics.

2.1 Acquisition of point clouds

Point clouds can be acquired through both direct and indirect methods, each with its own set of techniques depending on the application and the desired level of detail. These methods vary significantly in terms of accuracy, speed, and the types of environments they can handle, such as indoor versus outdoor scenes, static versus dynamic objects, or large-scale versus small-scale environments.

2.1.1 Direct acquisition

Point cloud data direct acquisition involves sensors or systems that capture 3D information in real-time or near real-time, creating a point cloud based on the physical environment. The resulting image can consist of sparse points in a 3D space, or of a dense cloud where each pixel on the sensor has an associated depth value, allowing a conversion into a dense point cloud.

2.1.1.1 Photogrammetry

In this method, 3D point clouds are created from RGB imagery by matching points across multiple images of the object or scene, all taken with different viewpoints [39]. Generally, a differentiating point in an image is identified and matched afterward to another view of the same scene. The features suitable for matching can be identified by SIFT [40] or other

algorithms for local feature detection [39]. Knowing the relative position of the cameras, the search for a point in the second image is restricted to a line by the epipolar constraint. Having multiple cameras can constrain the search zone even further [39].

2.1.1.2 Photogrammetry with structured light

This method projects a specific light pattern, such as grids, onto an object. Then, a camera captures how that pattern is deformed over the surface of the object. The deformation of the light patterns results on the depth information, which is then used to build the point cloud [41]. The determination of the surfaces 3D structure without features, such as walls and floors, is the main advantage over photogrammetry under ambient illumination. Photogrammetry without structured light tries to match the image point across different views and often fails when there is no sufficient visible texture on the image.

2.1.1.3 Light Detection and Ranging (LiDAR)

LiDAR emits laser pulses and measures the time it takes for the light to reflect back to the sensor [42]. This time-of-flight data is then used to calculate distances and generate a 3D representation of the scene. LiDAR is often used for large-scale mapping tasks, such as terrain mapping, autonomous vehicles, and archaeology [43]. LiDAR systems differ from laser-structured light systems because the distance is not measured by the time a single laser pulse takes to return. Instead, in a laser-structured light system, the distance is measured by the deformation of the appearance of an uninterrupted laser stripe or pattern caused by the 3D structure of objects [43].

2.1.1.4 Time-of-flight

The time-of-flight (ToF) camera is a range imaging camera system that measures distance based on the known speed of light, calculating the time of flight of the signal between the camera and the subject for each point of the image. Though the resolution for this kind of camera is usually low (sensor size ranging from 100 to 500 pixels), these systems are able to capture images in the order of tens to hundreds of images per second [44]. The two most common types, able to output a point cloud representation of the imaged scenes, are Photonic Mixed Devices and Range-Gated Images.

2.1.2 Indirect Acquisition

Indirect acquisition can be defined as algorithms applied to create point cloud data from sources that do not directly measure 3D information. Usually these models make use of

algorithms for the extraction of point cloud data from sets of 2D images or light field data. Sometimes, these algorithms extract dense depth data.

2.1.2.1 Structure From Motion

Structure from Motion (SfM) is a photogrammetry technique used to reconstruct three-dimensional structures from a series of two-dimensional images taken from different viewpoints [45]. It works by first capturing multiple overlapping images of an object or scene from various angles. Distinct features such as corners and edges are then detected in each image, and correspondences between these features are identified across the image set. Using these matched features, the relative positions and orientations (poses) of the cameras are estimated. With the camera poses established, a sparse 3D point cloud representing the scene is generated. This process essentially triangulates the positions of the matched features in 3D space. For more detailed reconstructions, additional points can be interpolated to create a denser 3D model. SfM is widely applied in fields such as computer vision, archaeology, and surveying for tasks like 3D modeling, mapping, and object reconstruction.

2.1.2.2 Light field data

The popularity of the light field cameras has been increasing in the last decade [46]. They are not only able to capture the light intensity in RGB spectral bands but also the intensity of the light leaving the scene, according to the angle at a particular point. This can be used to refocus, changing the viewpoint, depth of field, and extraction of 3D information from the image, all of which can be used for point cloud representation [47].

2.2 Point cloud coding

Point cloud codecs can be divided into three main categories, namely the geometric partition based (where octree is the most common), projection based and learning based.

2.2.1 Octree-based codecs

The most traditional point cloud coding models are based on the octree pruning method [48]. MPEG defined the Geometry-Based Point Cloud Compression (G-PCC) [1] for static point clouds. Fig. 2.1 shows the diagram for the G-PCC codec. The codec is based on an octree representation. Before encoding, the point cloud undergoes voxelization, a process that partitions the space into a regular grid of cubic units, referred to as voxels. Each point in the point cloud is subsequently assigned to the voxel that contains it.

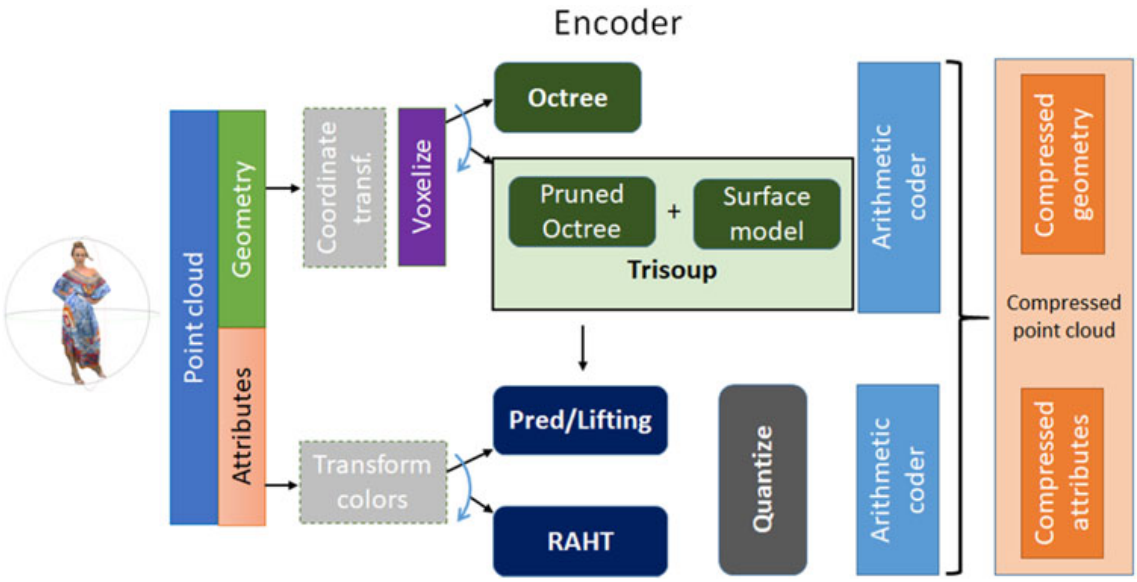


Figure 2.1: G-PCC reference encoder diagram (Figure taken from [1])

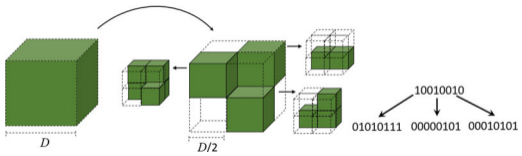


Figure 2.2: Example of octree decomposition (Figure taken from [1]).

Assuming the point cloud is limited to a quantized volume of $D \times D \times D$ voxels, the point cloud is divided vertically and horizontally into eight sub-cubes with dimensions $D/2 \times D/2 \times D/2$, as shown in Fig. 2.2. This is repeated until $D = 1$. For each step, occupied blocks are marked with 1 and unoccupied blocks are marked with 0. The generated octets are then compressed by an entropy coder that considers correlation with neighboring octets.

G-PCC also allows the alternative of using the trisoup method, which is based on surface reconstruction for geometry compression. In this case the geometry is represented by a pruned octree, which is constructed from the root to an arbitrary level. The leaves represent occupied sub-blocks that are larger than a voxel. Then, the surface of the object is estimated by a series of triangles. Since no connectivity exists between multiple triangles, the method is named triangle soup, or trisoup.

Furthermore, the point cloud attributes are compressed using either Region Adaptive Hierarchical Transform (RAHT) [49], predicting transform or lifting transform [1]. Since the lifting transform is built on the predicting transform, the merging of both is referred to as predicting/lifting transform. RAHT considers an octree representation of the point cloud, and starts from the highest level of the octree to the lowest. The transform is applied to each node, and performed for the x, y, z axis. The Predicting Transform is a distance-oriented predictive methodology for attribute encoding. This method uses the Level of

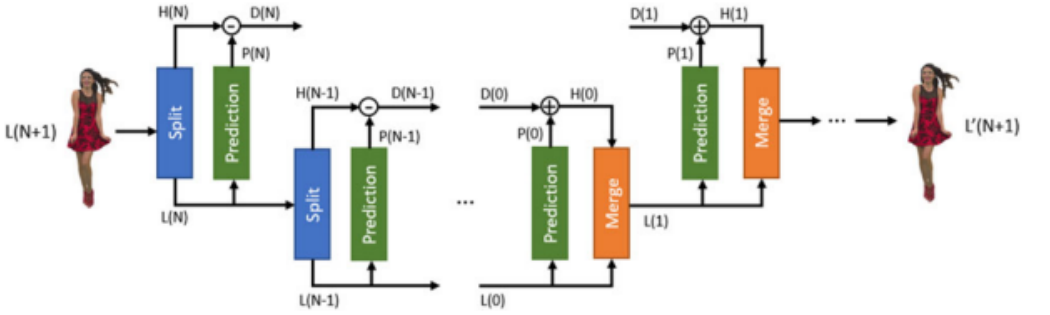


Figure 2.3: Illustrative diagram of Predicting transform (Figure taken from [1])

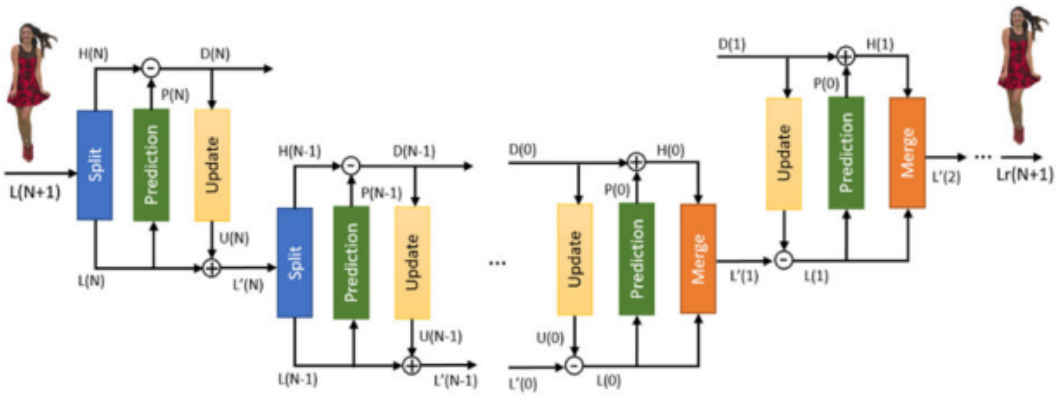


Figure 2.4: Scheme of the predicting/lifting transform (Figure taken from [1]).

Detail (LoD) representation, which distributes input points in sets of refinement levels (R), using the Euclidean Distance. The attributes of each point are then coded using the prediction determined by the LoD. The predicting transform is implemented using two operators based on the LoD structure, which are the split and merge operators. Assuming $L(j)$ and $H(j)$ as the sets of attributes associated with LoD(j) and $R(j)$, respectively, the split operator receives $L(j+1)$ as input, returning the low-resolution samples $L(j)$, as well as high-resolution samples $H(j)$. On the other hand, the merge operator takes both $L(j)$ and $H(j)$, returning $L(j+1)$. The process goes on recursively, and the reconstructed attributes are obtained through the cascade of merge operations. Fig. 2.3 illustrates the predicting transform. The lifting transform (Fig. 2.4) introduces an update operator, as well as an adaptive quantization strategy. In the LoD prediction scheme, each point is associated with an influence weight. Points in lower LoDs are used more often and, therefore, impact the encoding process more significantly. The update operator determines $U(j)$ based on the residual $D(j)$ and then updates the value of $L(j)$ using $U(j)$. The update signal $U(j)$ is a function of the residual $D(j)$, the distances between the predicted point and its neighbors, and their corresponding weights. Finally, to guide the quantization processes, the transformed coefficients associated with each point are multiplied by the square root of their respective weights.

2.2.2 Projection-based codecs

Another approach to point cloud coding relies on encoding the point cloud projections (an example is shown in Fig. 2.7), which can be coded using image or video codecs. MPEG also explored this approach, resulting in the Video-Based Point Cloud Compression (V-PCC) [1] for dynamic point clouds. The architecture of V-PCC is shown in Fig. 2.5.

V-PCC starts by projecting the point cloud into 3D patches. Patches are 3D surface segments generated by dividing the input point cloud into a number of connected regions. To generate the patches, the normal vector information for each point is estimated. Each orthographic projection direction is considered, and the points are then associated with the direction that yields the largest dot product between the normal vector of the point and the corresponding projection direction. Following the point classification process, the points that contain the same projection direction are connected, with each connected component being referred to as a patch. The points of each patch are projected considering the orthogonal projection to one of the six faces of the bounding box. This process is shown in Fig. 2.6.

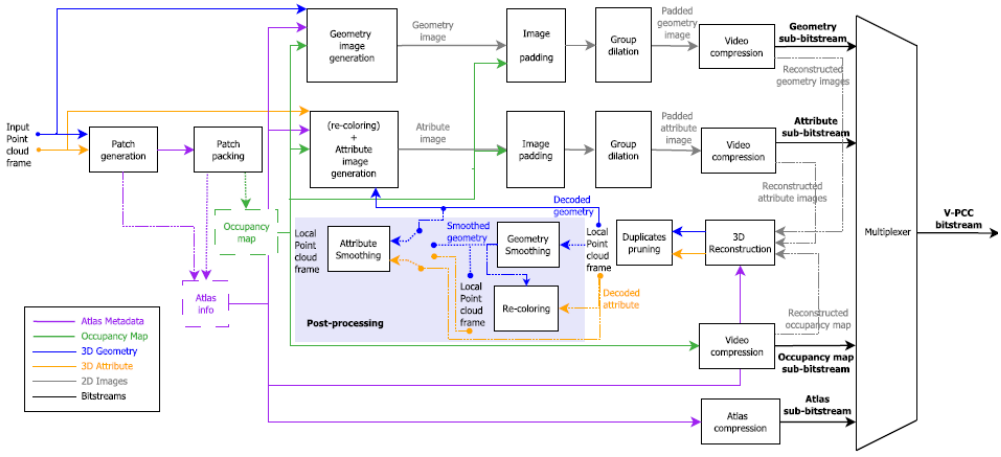


Figure 2.5: Architecture of V-PCC (Figure taken from [1]).

To generate images suitable for video coding, filtering operations need to be employed. One example is the definition of block sizes, in which the depth values cannot have larger variance, and points above a certain threshold are removed. It also defines a range to represent depth, and if the value is larger than the chosen range, the point is removed from the patch.

The projected patches are placed in an image of size $W \times H$, adjusted according to the intrinsic resolution of the point cloud being coded. The patches are ordered by size. The location of each patch is determined by searching in raster scan order, with the first location being selected. Blocks that contain pixels with valid depth values belonging to the area and covered by the patch size are considered occupied blocks and cannot be used by

other patches. This process makes sure that every block is associated with only one unique patch. If there is not available space for a given patch, the height of the image is doubled, and the positioning for that patch is re-evaluated. When all patches are inserted, the final image height is trimmed to the minimum necessary value.

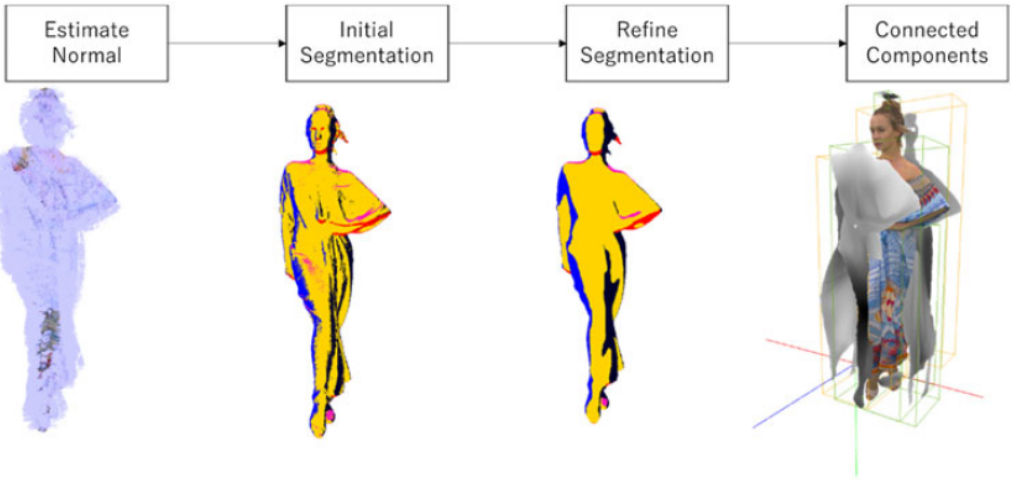


Figure 2.6: Patch generation process.(Figure taken from [1]).

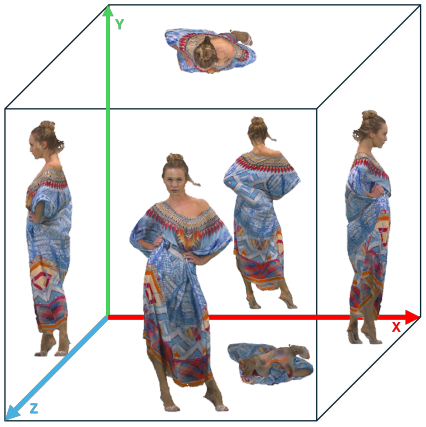


Figure 2.7: V-PCC projection into bounding boxes.

In V-PCC, geometry images capture the gap between the missing coordinates of point positions and the projection surface in the 3D bounding box, using only the luminance channel.

Due to the random shapes of patches, certain pixels may remain unoccupied following patch packing. To distinguishably assess the unused pixels and the pixels that will be considered to reconstruct the geometry information, an occupancy map is generated. This is a binary map, with a precision of $B \times B$ blocks, that are defined by the user. Pixels with

a value of 1 mean that at least one valid pixel is present in the corresponding $B \times B$ block, while the value of 0 indicates an empty space that was filled with pixels from the image padding procedure.

The resulting image sequences are then encoded using a video codec. In this procedure, when the resolution is reduced, it is required that the occupancy map image be scaled to a nominal resolution. This process can sometimes lead to inaccuracies in the occupancy map, resulting in the end effect of adding extra points to the reconstructed point cloud.

For geometry image padding, V-PCC uses a padding function to occupy the empty space between patches. The padding function aims to generate a smooth image suited for video compression. Each block of $T \times T$ pixels is processed independently. If the block is not occupied, the pixels of that block are filled with the last row or column of the previous $T \times T$ block, in raster scan order. This padding process is not necessary if the block is fully occupied. It can occur that the block contains both valid and non-valid pixels. In this case, the empty positions are sequentially filled with the values of the non-empty neighbors. This procedure is performed for each frame.

The raw format of YUV420, with 8-bit luminance and subsampled chroma channels, is used as input for video coding. The geometry information is packed in the luminance channel, as the geometry is represented by the distance, and contains only one component.

As the reconstructed geometry is likely to be different than the undistorted one, the color from the reference point cloud is transferred to the reconstructed geometry. This process considers the color information from a neighborhood of the nearest point on the reference point cloud. After the color values are determined, the color is mapped from 3D to 2D using the same mapping applied to geometry. A mip-map interpolation procedure is conducted to pad the color image. This process generates a multi-resolution representation of the texture map based on the occupancy map. This preserves active pixels, even when they are down-sampled along with empty pixels. To fill empty pixels at each resolution scale, a Gauss-Seidel [1] relaxation is employed. To optimize the process at higher resolutions, the lower resolutions are used to set the initial values. The sequence of padded images is converted to the YUV420 color space and coded with traditional video encoders.

The reconstruction process uses the decoded bitstreams for the occupancy map, geometry, and attribute images to reconstruct the 3D point cloud. When the near and far maps are used and the values from the two depth images are the same, several duplicate points may be generated by the codec. This is undesired, as they can have an impact on quality. To avoid this, the reconstruction process generates one point per (u, v) coordinate of the patch.

The coding of geometry and attribute images, along with the additional points introduced by occupancy map sub-sampling, generates artifacts that affect the reconstructed point cloud. Such artifacts can be reduced by point smoothing. A commonly used method iden-

tifies the points located on the edge of the patches and computes the centroid of the decoded points on a small $2 \times 2 \times 2$ grid. Then, a trilinear filter is applied.

To assess color related artifacts, the codec employs attribute smoothing techniques to reduce the seam effects on the reconstructed point cloud.

2.2.3 Learning based Codecs

Following the good performance in image coding [50], several learning-based coding solutions for point clouds have been proposed recently [8, 51, 52, 6, 7, 53, 54]. Learning-based encoders usually cause distortions that are quite different from those caused by conventional codecs. Typically, the point clouds are divided into blocks. During the decoding process, certain blocks may fail to be reconstructed, resulting in the appearance of empty spaces in the point cloud geometry.

These types of codecs are becoming more and more common, mainly due to the popularity of deep-learning technology [55]. Because of their rising popularity, it is crucial that objective quality models are able to accurately benchmark these kinds of solutions.

In the literature, there are codecs that can encode only geometry [52, 6, 7], only attributes[56, 57, 58, 59, 60, 61], and both color and geometry [8, 30]. Here, a brief explanation of the most relevant solutions will be provided.

2.2.3.1 Geometry only

Multiscale Point Cloud Compression (PCGCv2) [6] performs block-wise multi-resolution encoding. The point cloud is downsampled three times, and the encoding is done recurring to the Inception Residual Network [62]. At the bottleneck, the geometry coordinates are encoded with G-PCC, and entropy coding is used for the attributes. The decoding branch architecture mirrors the encoder. The architecture is shown in Fig. 2.8. In the figures, conv $c \times n^3$ denotes the sparse convolution operation with c output channels, as well as an $n \times n \times n$ kernel size. The upscaling and downscaling, with a given factor of s, are shown by $s \uparrow$ and $s \downarrow$. The arithmetic encoder and decoder are depicted by AE and AD.

The model is trained with densely sampled data from the ShapeNet database [63]. The final training set is obtained by random rotation and quantization with 7-bit precision and a randomized number of points.

Different coding bitrates are targeted by varying the rate-distortion tradeoff parameter λ between 0.75 and 16. The code made available defines the global loss function as $J = \alpha D + \beta R$ where D is the distortion and R is the coding bitrate.

Deep Point Cloud Geometry Compression (GeoCNNv2) [7] learns an encoding function

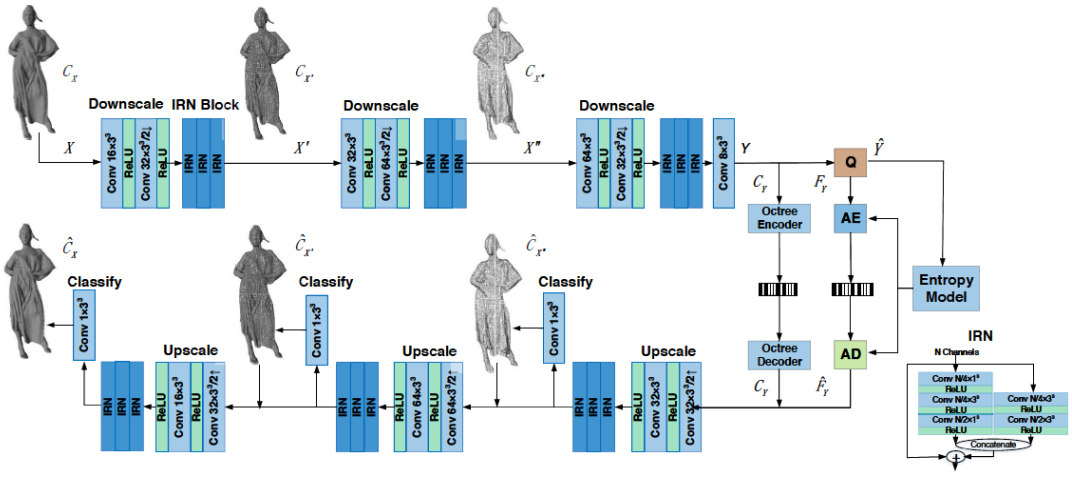


Figure 2.8: Architecture of PCGCv2, and IRN block. (Figure taken from [6]).

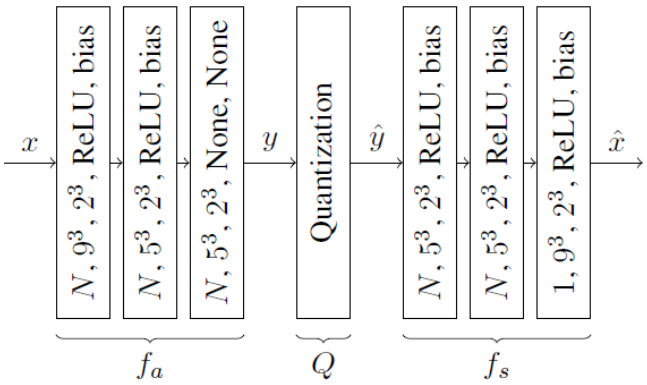


Figure 2.9: Architecture of GeoCNNv2. (Figure taken from [7]).

from three sequential convolution layers. The first two use ReLU activation. The latent representation of the third label is quantized through element-wise integer rounding and then compressed through a combination of algorithms. The architecture of the codec is shown in Fig. 2.9. The decoding architecture mirrors the encoding. The output of the last layer is converted to the distorted point cloud using element-wise minimum, maximum, and rounding functions.

PCC GEO CNNv2 trains four individual models for each Rate-Distortion tradeoff given by $J = \lambda D + R$ [7]. They chose four values for λ, notably 3×10^{-4}, 10^{-4}, 5×10^{-5}, 2×10^{-5} and $\lambda = 10^{-5}$. It follows a sequential training approach, with successively decreasing values of λ. The trained weights for λ_{i-1} are used to initialize the training for λ_i).

The models were trained on a subset of the ModelNet40 [64] dataset. First, the mesh data is voxelized with a resolution of $512 \times 512 \times 512$, and the 200 largest point clouds are selected. Then, the point clouds are divided into blocks with a resolution of $64 \times 64 \times 64$, and the 4000 largest blocks are selected. For each value of λ, the model was trained for

500 steps, with early stopping if the loss did not improve for more than 4 validation steps.

ADLPCC [53] partitions the point cloud into regular-sized 3D blocks. Several models separately code those blocks. The codec contains an autoencoder (AE) and a variational autoencoder (VAE) with three convolutional layers of both encoding and reconstruction, with sigmoid and ReLU activations, respectively.

The global loss function of ADLPCC[1] is given by $J = D + \lambda R$, where the coding rate R is estimated during training as the summed entropy of its autoencoder and variational autoencoder latent representations.

In order to obtain several Rate-Distortion tradeoff points, different λ values are considered, thus varying the weight of the rate. The model is trained with a dataset consisting of point clouds collected by JPEG and MPEG [53], with $\lambda = \{500, 900, 1500, 5000, 20000\}$. For each value of λ, the codec is trained with $\alpha = \{0.5, 0.6, 0.7, 0.8, 0.9\}$, which is a parameter of the BCE focal loss function [53], which allows choosing the best performing model considering the characteristics of the point cloud, such as its sparsity.

The Resolution Scalable Deep Learning Point Cloud Compression (RSDLPCC) [52] makes use of deep learning technology to compress the point cloud geometry. A latent representation of a point cloud is computed by an autoencoder framework. The interlaced block creation makes the scalability feature possible. The point cloud is divided into superblocks that are further divided by interlaced downsampling. This procedure creates eight interlaced blocks for each defined superblock. The resulting blocks are then coded separately, enabling random access. The training process is based on ADLPCC [53].

2.2.3.2 Color Only

Deep-learning technology has also been researched to encode point cloud attributes. Usually, one of the two methodologies is followed. Either the point cloud attributes are mapped onto a 2D image and then are coded with traditional image codecs, such as VVC, HEVC, or JPEG AI, or they are coded by using point convolutions.

Quach *et al.* [65] proposed a folding based approach for point cloud compression. It starts mapping the point cloud by training a deep neural network to fold a 2D grid onto a point cloud. The attributes of the point cloud are then mapped onto the grid, resulting in an image that can be coded with BPG[2], a format based on HEVC [66]. Alexiou *et al* [67] considered the convolutional neural networks used for geometry compression in order to code attributes. Sheng *et al.* developed a point based neural network for attribute coding [68].

TSC-PCAC [69] includes a framework consisting of a Transformer and Sparse Convolution Module (TSCM) based variational encoder and channel context module. The TSCM

[1] https://github.com/aguarda/ADLPCC
[2] https://bellard.org/bpg/

consists of two stages. The first stage modules local dependencies and feature representations of the input point cloud. The second stage captures global features through spatial and channel pooling encompassing.

Recently, the ScalablePCAC codec was proposed [57]. G-PCC is used at the base layer to encode the thumbnail point cloud, which is downscaled from the original input. A learning-based model is implemented at the enhancement layer to compress and restore the full-resolution point cloud conditioned on the base layer reconstruction. A cross-layer rate allocation strategy determines the resolution downscaling factor, the quantization parameter, and the quality controlling factor of the enhancement layer.

3DAC [70] was proposed as another method to encode point cloud attributes, such as color and reflectance.

They are, firstly, converted to transform coefficients. A deep entropy model then models the probabilities of these coefficients by considering information hidden in attribute transforms and previously encoded attributes. The estimated probabilities are used to further compress these transform coefficients to a final attributes bitstream.

2.2.3.3 Joint Geometry and Color - The JPEG Pleno Learning Based Point Cloud Coding Standard

In response to the JPEG Pleno Point Cloud Coding Call for Proposals [29], the codec submitted by the IT/IST team [8], named IT-DL-PCC, was selected as the best performing one. This solution has been selected to be the starting point of the standard under development by the JPEG Pleno Point Cloud Coding Committee [9].

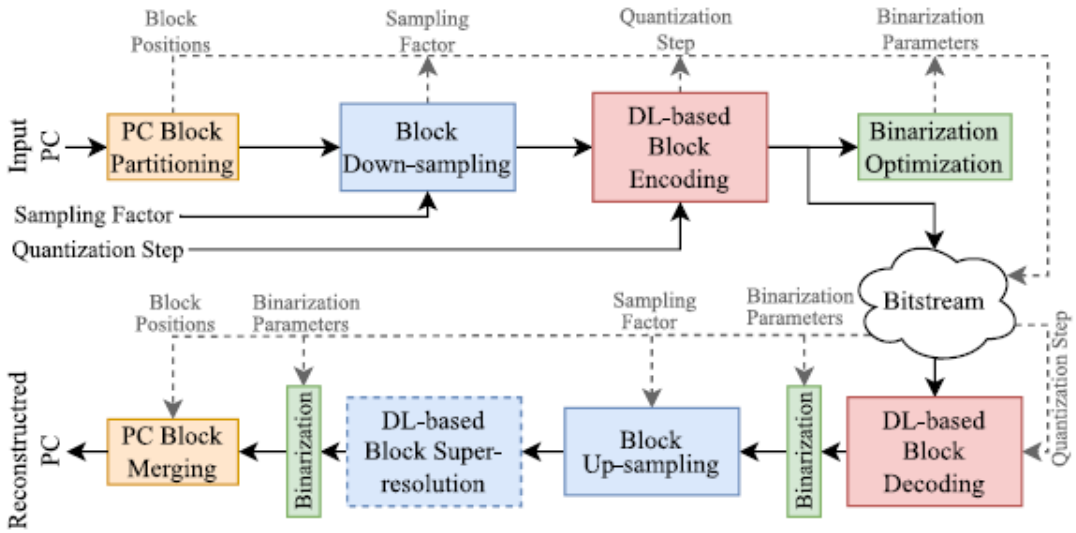


Figure 2.10: Overall IT-DL-PCC architecture (Figure taken from [8]).

Fig. 2.10 shows the general architecture of the codec.

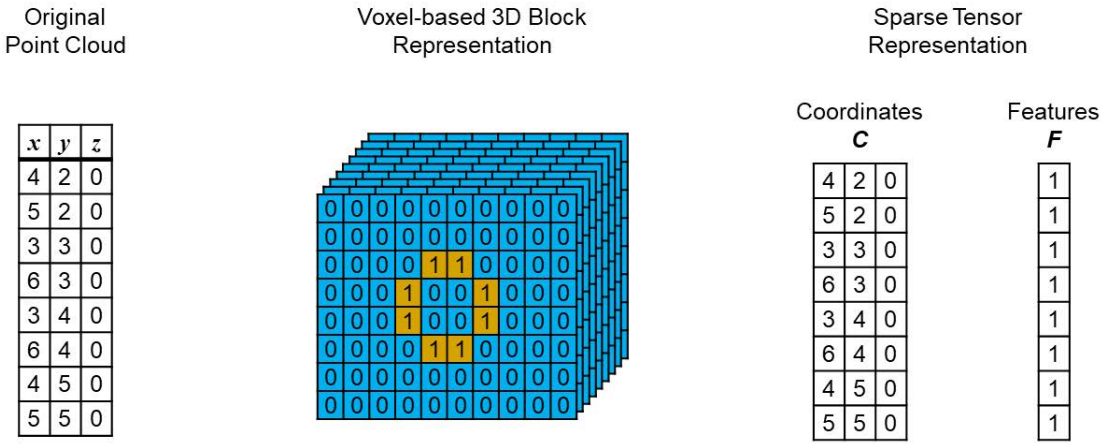


Figure 2.11: Example of sparse tensor representation (Figure taken from [9]).

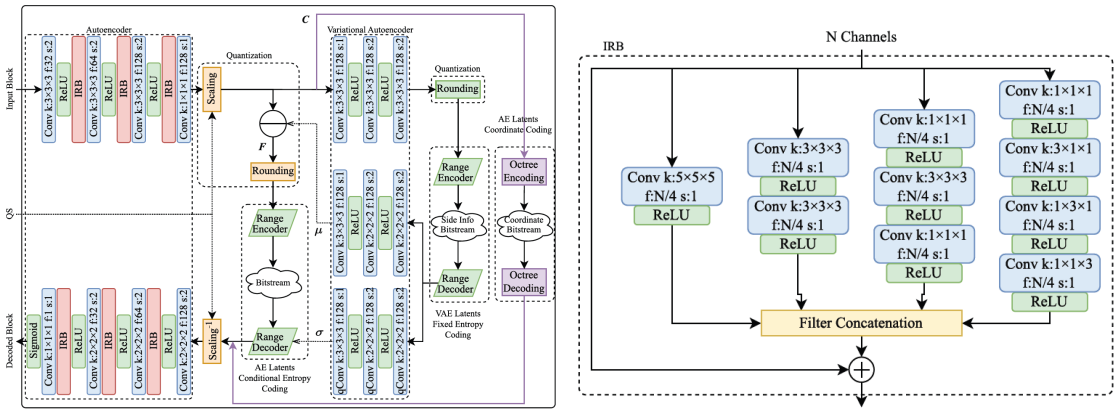
Prior to the encoding process, the point cloud is converted to a voxel-based 3D block representation. This allows to represent the point cloud in a regular structure, suitable for the usage of convolutional neural networks (CNN). The geometry data is represented as a binary signal, where 1 corresponds to an occupied voxel. Conversely, 0 represents an empty voxel. However, this representation is very computationally expensive. Instead, a sparse tensor representation is considered, as it only requires representing the non-empty voxels by their respective coordinates and corresponding features, while the remaining voxels are assumed to be empty [9]. An example is shown in Fig. 2.11.

Given this representation, the point cloud is reorganized by dividing it into blocks of a specific size that can be coded separately with a DL model. The position of each block is transmitted to the decoder. At the decoder side, the full point cloud is reconstructed by merging the blocks according to their position.

Depending on the characteristics of the point cloud, block down-sampling or up-sampling is employed in order to reduce the point cloud coding precision at the encoder side, achieving a more efficient compression at the encoder side. At the decoder side, these steps allow the point cloud to be restored to the original precision. To achieve this, the input point cloud is scaled by a sampling factor, determined by the user, followed by a rounding operation. This results in a loss of points and a denser surface. At the decoder, the coordinates of the reconstructed point cloud are scaled back using the inverse sampling factor.

The end-to-end geometry coding model is shown in Fig. 2.12. All the convolution layers in the image are sparse convolutions, implemented with Minkowski Engine v0.5.4 [71].

The generated blocks are the input of the autoencoder, which transforms them into a latent representation. This is compared to the transform coding stage, usually present in traditional image coding. The latent representation consists of multiple feature maps, and their number depends on the selected number of filters for the convolutional layers. The autoencoder is a combination of 3D convolutional layers, and Inception-Residual Blocks (IRB). The IRB is built upon the Inception ResNet [62], containing several convolutional



(a) Overall Geometry Coding Model

(b) Inception-Residual Blocks (IRB)

Figure 2.12: End-to-end DL geometry coding model architecture. (Figures taken from [9]).

layers and different filter sizes. This allows to extract different kinds of features from different neighboring contexts. A residual skip connection enables the propagation of features along the network, making the training of deeper models easier. Initially, the number of filters is set to 32, and is progressively increased to 128 at the final layer, providing a rich latent representation.

The coordinates (C) present in the latent representation generated by the AE are losslessly coded using G-PCC [1].

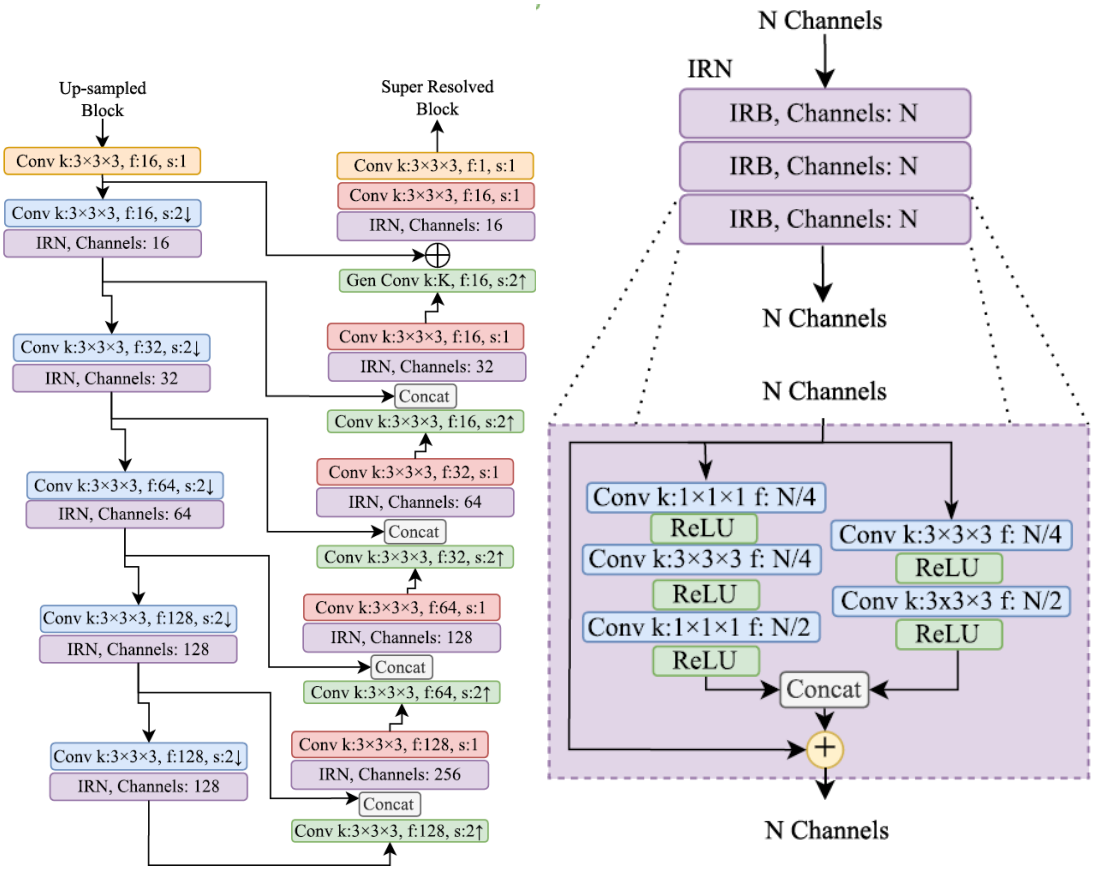
A variational autoencoder (VAE) captures structure information that may be present in the block latent representation. It is then used as a hyperprior for the conditional entropy coding model. In the case of the VM, the mean scale hyperprior is used [72]. The VAE generates a latent representation that must be coded and transmitted in the bitstream to the decoder. The VAE consists of three convolutional layers at the encoder side, as well as a symmetric decoder with 2 identical branches. The first produces the means μ, while the second one generates the scales σ (standard deviations).

The latents of the VAE are coded using an entropy coding model for all blocks, which is learned during training.

Before entropy coding, the latent representation of the AE is explicitly quantized. The latents are scaled by a user-defined quantization step (QS), which must be a positive real value. Then, the means μ generated by the VAE are subtracted, achieving a residue latent representation, rounded to the closest integer. This approach allows to tune the target rate at coding time for an individual model.

The features (F) present in the resulting residue latent representation are coded with a conditional entropy coding approach. A Gaussian scale mixture is considered, conditioned on a hyperprior as the entropy coding model. The obtained scales σ are also considered. During the training process, the entropy of the latent representation is evaluated according to the entropy coding model, used for the rate-distortion optimization process.

At the decoder side, each block is decoded with the same architecture shown in Fig. 2.12. The ”Side Info Bitstream” that contains metadata related to entropy coding is decoded. This generates the entropy coding model parameters used for the current clock before decoding the final bitstream [9].



(a) Geometry SR Model (b) Detailed IRN Block

Figure 2.13: End-to-end DL geometry coding model architecture. (Figures taken from [9]).

The model also contains an optional DL-based block super-resolution model [9]. The model receives the output of the block up-sampling module, meaning the point cloud is at the original precision. However, it is now sparser. The purpose of this model is to densify the point cloud. This may increase quality at no rate cost, although it requires additional computational complexity.

However, this is not always the case, namely if the point cloud is sparse or if it contains many coding artifacts or not. Considering this, the super-resolution model was implemented as an optional post-processing module.

The architecture is shown in Fig. 2.13. It consists of a 3D CNN shaped as a U-net [73]. All the convolution layers in the image are sparse convolutions, implemented with Minkowski Engine v0.5.4 [71].

The first path of the U-net extracts features at different scales.

It combines 3D down-sampling convolutional layers and IRB blocks. The IRB blocks are much simpler and lighter, as they have fewer and smaller filters. On the contrary, the geometry SR model contains many more convolutional layers and IRBs. The number of filters starts at 16 in the initial layer, progressively increasing until 128.

The second path up-samples the features, with the task of also aggregating the multiscale features extracted in the contracting path. By considering this, the model is able to predict the occupation of the voxels lost due to the downsampling process.

The output values of the geometry coding model are in the range of [0, 1], for each block. This represents the probability of a voxel being occupied. Consequently, their probabilities need to be transformed into binary values, which will correspond to the final reconstructed points. A *Top-K* binarization method is used in the codec, to select the voxels that are occupied. In this method, only the *k* voxels with the largest probabilities are selected as occupied, as defined in eq. 2.1,

$$k_{codec} = N_{input} \times \beta \quad (2.1)$$

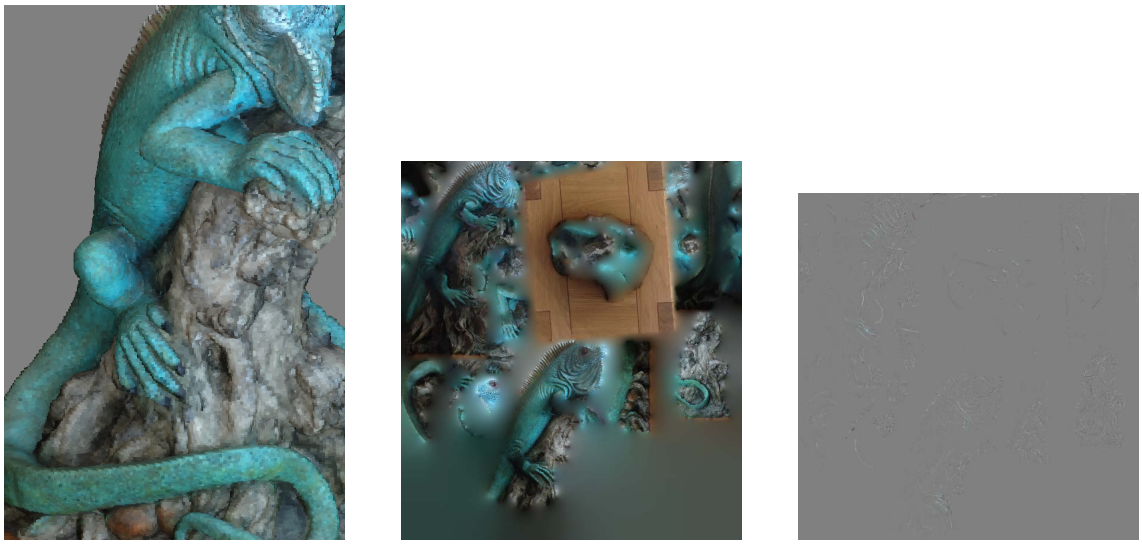
where N_{input} is the number of known points in the original block, and β is a factor optimized at the encoder, using linear search algorithms. The value that results in the best reconstructed quality is selected. The output of the super resolution model also requires this process, as the output is also binary.

The color attributes information is coded after the geometry information. The geometry coding process is lossy and leads to the points being in different positions when compared to the reference point cloud. A recoloring process is used to transfer the color information from the reference point cloud to the decoded geometry. Two different steps exist in this process:

- For points with a direct correspondence, the color of the reference point cloud is simply copied.
- When there is no direct correspondence, the color is interpolated using a Radial Based Function (RBF) interpolation, with a linear kernel. This approach considers the 20 nearest neighbors.

Then, the point cloud projections extracted using the V-PCC projection model [1]. To extract the 2D projections, the V-PCC standard is used. It was slightly modified to remove the color dependence during the path refinement stage. This is conducted to eliminate the need for additional information at the decoder side when an inverse projection is conducted.

V-PCC generates two different maps, namely the near map and the far map. They can be very similar, depending on the complexity of the point cloud geometry.



(a) Reference Point Cloud (b) Near Layer (c) Far Layer

Figure 2.14: V-PCC projection example for the Iguana PC. (Images 2.14b and 2.14c taken from [9]).

To avoid this, the difference between the far and near layers is computed. The far layer residue image is trimmed, considering the smallest rectangle totally filled with non-zero residue pixels. The image is then represented as 8-bit unsigned integers. This is done by scaling the image by a factor of 2 and a shift of 128, allowing it to reach the range of [0, 255]. Fig. 2.14 shows an example of a projection obtained using the V-PCC standard.

The resulting projections are encoded with JPEG AI [50, 74]. The color super-resolution model is used to determine the new points generated by the geometry super resolution model. The color of the new points is computed using the RBF interpolation with linear kernels and considers the 20 nearest neighbors. The geometry coding model is trained by minimizing a loss function (LF) that considers the distortion of (D) each coded block and the estimated coding rate (R), using a Lagrangian multiplier λ given by equation 2.2:

$$LF = D + \lambda \times R \quad \textbf{(2.2)}$$

A different model is required for each RD point. A total of five models are considered in the standard, namely $\lambda = 0.0025, 0.005, 0.01, 0.025, and 0.05$. The models are trained sequentially, from the smallest to the largest [7]. The first model is trained without any particular initialization. The remaining ones are initialized with the weights obtained by the previous models.

Since binary optimization is required at both the encoder and the decoder step, it cannot be performed during training, as it is not a differentiable operations. To solve this, the geometry distortion is measured as the average voxel level distortion. It is computed as a

binary classification using a the Focal loss (FL), defined in eq. 2.3:

$$FL(u,v) = \begin{cases} -\alpha(1-v)^{\gamma}\log v, & u=1 \\ -(1-\alpha)v^{\gamma}\log(1-v), & u=0 \end{cases} \quad (2.3)$$

where u is the binary value of the original voxel, and v is the corresponding probability value of the corresponding decoded voxel. The α parameter controls the class imbalance effect, and the parameter λ increases the importance of correction of misclassified voxels. These parameters are set to 0.5 and 2, respectively.

The learning rate is set to $10^{-}4$. If after ten epochs, the validation loss does not decrease below 10%, it is reduced to $10^{-}5$.

2.3 Subjective quality evaluation of Point Clouds

This section will cover the typical procedures followed in order to conduct subjective quality evaluation studies.

2.3.1 Historical Background of Subjective Quality Evaluation

Quality evaluation of images started with the emergence of photography in the 19th century. In those days, photographers and critics evaluated images based on facts like sharpness, contrast, and composition. With the rise of cinema in the late 19th and early 20th centuries, both filmmakers and audiences evaluated film quality based on subjective criteria, such as visual clarity or resolution.

With the creation of television in the 1930s came a necessity to evaluate the quality of broadcast images. Subjective quality protocols were developed to assess picture quality, focusing on aspects such as resolution, contrast, and the presence of virtual artifacts.

Given the exponential growth of television, it became necessary to standardize methods to evaluate image quality. Researchers developed subjective testing methodologies in which human observers would rate image quality based on controlled viewing conditions and standardized rating scales.

The International Telecommunication Union Radiocommunication Sector (ITU-R) played a key role in standardizing image quality evaluation methods. ITU-R Recommendation BT.500 [25], first published in 1974, contained a standardized methodology for subjective assessment of television picture quality. This recommendation has been in constant update to accommodate advancements in imaging technology. It also provides several methodologies for conducting subjective tests, including the selection of test images, viewing conditions, and rating scales, as well as statistical analysis of the results.

With the emergence of digital imaging, new challenges for subjective quality evaluation appear, such as evaluating compression artifacts, noise, and color fidelity.

To tackle this problem, new subjective quality evaluation protocols were developed for digital images, namely the Double Stimulus Continuous Quality Scale (DSCQS) and the Single Stimulus Continuous Quality Evaluation (SSCQE).

2.3.2 Related Work

Several studies were carried out to establish quality models for geometry-only [75, 76], graph-based [77], and projection-based [78] codecs. Honglei *et al.* [79] conducted subjective evaluations to study coding solutions that would become the early stages of MPEG standards, namely the codecs V-PCC and G-PCC. A study on those coding solutions before their final standardization was also reported [80].

Objective and subjective quality evaluations were conducted to assess the performance of MPEG standards using a 2D setup [81]. It was concluded that V-PCC was the best-performing codec. Those solutions were compared with Draco[3] and RSDLPCC [52] using a 2D display [16], followed by a study where the 2D display was replaced by a 3D stereoscopic visualization setup [14]. The previous study was expanded [5] in order to evaluate several point cloud metrics.

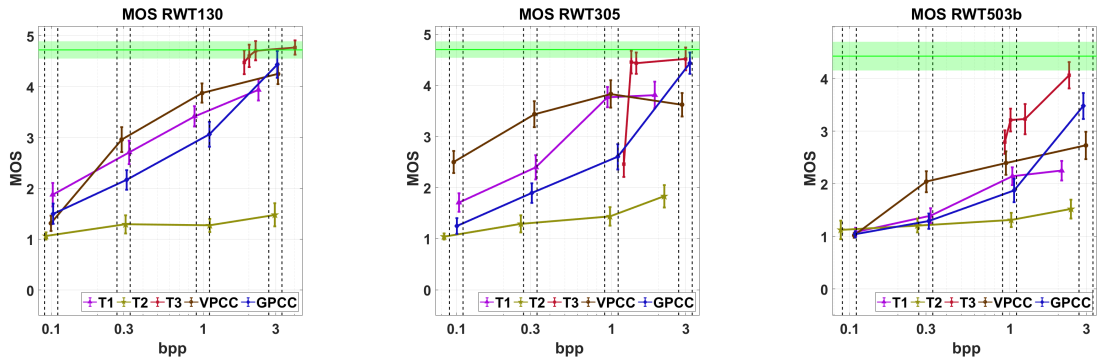
Subjective evaluation studies using virtual or augmented reality (VR/AR) environments have also been reported [82, 83, 84, 85, 86, 87]. Moreover, a subjective quality assessment study targeting learning-based coding solutions [13] was conducted, using a set of six point clouds depicting both objects and landscapes encoded with three deep-learning-based codecs. The tested learning-based solutions showed competitive results when compared with G-PCC.

JPEG Pleno Point Cloud Coding activity also reported a subjective study using state-of-the-art codecs prior to the launch of its call for proposals [88]. The aim was to evaluate state-of-the-art point cloud coding solutions, analyze the stability of subjective quality assessment models, and evaluate the performance of objective metrics. The call for proposals results were reported by Prazeres *et al.* [10]. Three deep-learning-based solutions capable of encoding both geometry and color were evaluated, and the best-performing solution became the base for the JPEG Pleno Point Cloud Learning-based Verification Model [30]. Subjective quality evaluation studies also lead to the definition of point cloud quality assessment databases, commonly used to benchmark point cloud quality metrics [89, 90, 91, 92]. Crowdsourcing methodologies have also been studied [18, 93] as a method of subjective evaluation. Lazzarotto *et al.* [94] reported an evaluation of different tradeoffs between the geometry and texture bitrate. Three different tradeoffs between geometry and texture were evaluated, and compared to assess which one would achieve

[3] https://github.com/google/draco

(a) *RWT130* (b) *RWT305* (c) *RWT503b*

Figure 2.15: Examples of point clouds used in the JPEG Pleno Point Cloud Call for Proposals responses evaluation(Figures taken from [10]).



(a) *RWT130* (b) *RWT305* (c) *RWT503b*

Figure 2.16: Example of MOS vs bpp with 95% confidence intervals for the point clouds of Fig. 2.15. The green bar in top represents the confidence interval obtained. (Plots taken from [10]).

the best performance. The main conclusion was that the best tradeoff is very dependent on the point cloud content.

Fig. 2.16 shows an example of the subjective analysis performed during the doctoral program [10], analyzing V-PCC, G-PCC, and the solutions submitted to the JPEG Pleno Point Cloud Coding Call for Proposals, named T1, T2, and T3. The y-axis represents the obtained MOS. The x-axis represents the bitrate in bits per point (bpp). Each line represents a codec. The vertical continuous lines represent the 95% confidence interval, assuming a Gaussian distribution. The green line represents the MOS obtained for each reference pair, and the green shade around it represents the 95% confidence interval. The vertical dashed lines represent the tolerance interval for each target bitrate (in this case,a bit rate deviation of $\pm$ 10% was accepted).

2.4 Objective quality evaluation

Lossy point cloud coding solutions might induce significant visual distortions that need to be accurately measured by reliable models that evaluate the quality of the decoded point cloud content. The most reliable quality measures result from suitable subjective quality tests that require careful planning and are very time-consuming. Instead, developers consider objective quality measures that allow the quality evaluation of point cloud compression methodologies.

Objective quality evaluation metrics are crucial in the development of coding methods, as they do not require long and costly subjective quality evaluation. These metrics are generally developed to most accurately reflect subjective quality. Some metrics only represent a measure of the signal fidelity, such as PSNR MSE D1 [19], but in those cases, the representation of the subjective results is frequently limited. Other metrics try to simulate the human perception of quality [11, 37] This section analyzes the state-of-the-art in the objective quality assessment of point clouds.

Point cloud quality evaluation metrics can be divided according to the type of information under evaluation:

1. Geometry only: Only considers the geometry of the point cloud,
2. Color only: Only consider the color attributes of the point cloud, and
3. Geometry and color/luminance: Considers both the geometry and color attributes of the point cloud.

Furthermore, they can subdivide into three major groups, considering the type of data they use:

1. Full-reference metrics: Where the distorted data is compared with the original data
2. Reduced-Reference: where a set of features extracted from the distorted and reference point clouds are compared.
3. No-reference: Where only the distorted point cloud is considered

2.4.1 Full Reference Metrics

2.4.1.1 PSNR MSE D1 [19]

This metric is based on the geometric distance of points between the reference and the distorted point cloud. For every point b_k in the distorted point cloud, a corresponding point a_i in the reference is identified through the nearest neighbor algorithm. Then, the

individual error is computed using the Euclidean distance between two points $E(a_i, b_k)$, using equation 2.4.

$$E(a_i, b_k) = |(\overrightarrow{v}_{b_k}^{a_i})|_2 \quad (2.4)$$

The process is repeated for every point b_k, thus indicating the displacement between the reference point cloud and the distorted point cloud [19], and the average of individual the errors are then computed. The reference and distorted point cloud are swapped in equation 2.4 and a new average of individual errors is computed. The final value of the metrics is given by the lowest value of the two averages.

2.4.1.2 PSNR MSE D2 [19]

This metric requires the normal vectors of the reference point cloud. For every point b_k of the distorted content, a point a_i in the reference content is identified, and then the projected error $E(a_i, b_k)$ across the normal $N_(a_i)$ of the corresponding reference point is computed, using equation (2.5):

$$E(a_i, b_k) = \left|(\overrightarrow{v}_{b_k}^{a_i} \cdot (N_{a_i})\right| \quad (2.5)$$

Similar to PSNR MSE D1, the reference and distorted point clouds are swapped in equation 2.5, and a new average of individual errors is computed. The final value of the metrics is given by the lowest value of the two averages.

2.4.1.3 PSNR MSE YUV [95]

This metric is based on the error of the color values between the identified point in the reference and the distorted point cloud. The identification process is conducted using the nearest neighbor algorithm. An individual error is computed for the identified points based on the Euclidean distance. For color attributes, the mean squared error (MSE) or Hausdorff distance (HAU) is calculated for the three components, with an RGB to YCbCr conversion being made [96]. The PSNR value is computed using the equation (2.6):

$$PSNR = 10\left(\frac{peak^2}{MSE, HAU}\right) \quad (2.6)$$

with *peak* being 255, considering that all the color attributes of the tested point clouds have a bit depth of 8. The metric is then computed symmetrically. The final value for each color channel is the maximum between the two computations. The final value for

the metric is the PSNR value for each color channel, computed by equation (2.7).

$$PSNR_{Colour} = \frac{6PSNR_Y + PSNR_{C_b} + PSNR_{C_r}}{8} \quad (2.7)$$

2.4.1.4 Point Cloud Quality Metric (PCQM) [11]

PCQM considers a data driven approach. Different features are selected and computed, and then the metric is computed as a linear combination of an optimal subject of the computed features. A correspondence is established between the distorted point cloud D and the reference cloud R, defining a neighborhood for each point. The correspondence is made by computing quadric surface fitting on a set of nearest neighbors $p_i^D \in D$ of $p \in R$. This neighborhood is considered to be spherical, given the geometric correspondence between R and D. For color correspondence between point clouds, for each point in the computed quadric surface, the color of the nearest neighbor on D is assigned.

The metric extracts geometry-based features, relying on *mean curvature* information to calculate curvature comparison, contrast, and structure. Afterwards, the color based features are computed. The RGB values are converted to the LAB2000HL perceptual color space. Then, lightness comparison, lightness contrast, lightness structure, chroma comparison, and hue comparison are computed for each point. Local values for the features are then aggregated by average pooling. For each point p, local features f_i^p are computed between $[0, 1]$. The global features are obtained using the following equation:

$$f_i = \frac{1}{|R|} \sum_{p \in R} f_i^p \quad (2.8)$$

Finally, those features are combined into a perceptual quality score using a linear model

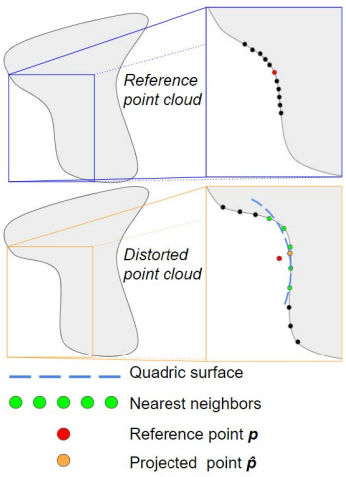


Figure 2.17: Illustration of the point-to-surface correspondence computation (Figure taken from [11]).

that was optimized through logistic regression.

$$\text{PCQM} = \sum_{i \in S} w_i f_i \quad (2.9)$$

where S is the set of indices of features in the linear model, f_i, the global features, and w_i, the weight attributed to those features.

2.4.1.5 Point Cloud Structural Similarity (PointSSIM) [97]

This metric extracts features to quantify the statistical dispersion of quantities that characterize the local topology and appearance of the point cloud. Neighbors around every point of a model are selected to capture local properties. Quantities to reflect local properties are computed, considering four different attributes, notably geometry information, normal vectors, curvature values, and texture information. For the feature extraction, dispersion statistics are computed using one of the available estimators, namely the median (m), variance (σ^2), mean absolute deviation (μ_{AD}), median absolute deviation (m_{AD}), coefficient of variation (COV), and quartile coefficient of dispersion (QCD). The estimators will be applied over a number of K nearest neighbors. Three different pooling methods are considered, namely the arithmetic mean (mean), mean square error (MSE), and root mean square (RMS). The perceptual quality prediction is based on the feature similarity values extracted from the reference point cloud (X) and the distorted point cloud (Y). Each neighborhood of Y is associated with a neighborhood of X, by identifying for every point *p* of Y the nearest point *q* in X. Then the similarity is measured as the relative difference between the corresponding feature values (F_X and F_Y), with ε being an arbitrary small value, in order to avoid undefined operations.

$$S_Y(p) = \frac{|F_X(q) - F_Y(p)|}{\max\{|F_X(q), F_Y(p)|\} + \varepsilon} \quad (2.10)$$

Finally, a final score S_Y for the model in evaluation is estimated through error pooling across all points, using the following equation:

$$S_Y = \frac{1}{N_p} \sum_{p=1}^{N_p} S_y(p)^k \quad (2.11)$$

with $k = \{1, 2\}$, denoting the mean and MSE, respectively.

2.4.1.6 GraphSIM [37] and MS-GraphSIM [12]

MS-GraphSIM [12] extends the GraphSIM metric [37] by computing its graph-based features at different scales to better represent the characteristics of the Human Visualization System (HVS). The metric starts by extracting a set of keypoints, $\overrightarrow{s}$, obtained by resam-

Figure 2.18: Example of multi-scale operations employed in MS-GraphSIM, namely low-pass filtering, downsampling and region shrinking. (Figure taken from [12]).

pling the reference point cloud ($\overrightarrow{P}_r$) geometry using a high-pass graph filter. The resulting point cloud, $\overrightarrow{P}_s$, shows high spatial-frequency regions like edges or contours. For both the reference and distorted ($\overrightarrow{P}_d$) point clouds, local graphs are constructed with each obtained keypoint $\overrightarrow{s}_k$ as its center. The neighbors of $\overrightarrow{s}_k$ are clustered using the Euclidean distance of the geometry components of corresponding points in $\overrightarrow{P}_r$ and $\overrightarrow{P}_d$. After constructing a local graph, the color information of the neighbors belonging to that graph is set as a graph signal and passed through a low-pass graph filter, amplifying the low frequencies and attenuating the high frequencies. The neighborhood undergoes down-sampling using systematic sampling [12]. The sampled points are moved towards the centroid of the point cloud bounding box. For each scale, three similarity measures are computed based on color gradient features, notably the gradient mass (m_g), the gradient mean (μ_g) and the gradient variance σ_g^2 and covariance (c_g). These features reflect the spatial variation of point density, the distortion of the statistical characteristics of the signals, and the spatial disturbance of the points, respectively, and are defined as follows:

$$m_g = \sum_{\overrightarrow{X}_j \in \mathcal{N}_k} \sqrt{W_{\overrightarrow{X}_j, \overrightarrow{s}_k}} [f(\overrightarrow{X}_j) - f(\overrightarrow{s}_k)] \quad (2.12)$$

$$\mu_g = \frac{1}{N} (\nabla_{\overrightarrow{s}_k} f) \quad (2.13)$$

$$\sigma_g^2 = \frac{\sum(g_j - \overline{g})}{N} \quad (2.14)$$

$$c_g = E[\overrightarrow{g}_{\overrightarrow{s}_k} \cdot \overline{g}_{\overrightarrow{s}_k}] - E[\overrightarrow{g}_{\overrightarrow{s}_k}] \cdot E[\overline{g}_{\overrightarrow{s}_k}] \quad (2.15)$$

In equation (2.12), $W_{\overrightarrow{X}_j, \overrightarrow{s}_k}$ represents the euclidean-based graph weight, $\overrightarrow{X}_j$ represents the color attributes of the point cloud, and $f(\overrightarrow{X}_j) - f(\overrightarrow{s}_k)$ is the attribute gradient. $\mathcal{N}_k$ is the set of points effectively connected to $\overrightarrow{s}_k$. In equation (2.13), N represents the number of points in $\mathcal{N}_k$. In equation (2.14), g_j represents the j-th element in $\overrightarrow{g}_{\overrightarrow{s}_k}$ and $\overline{g}$ is the weighted average gradient of $\overrightarrow{g}_{\overrightarrow{s}_k}$. Finally, in equation (2.15), $\overrightarrow{g}_{\overrightarrow{s}_k}$ and $\overline{g}_{\overrightarrow{s}_k}$ represent the weighted gradient distribution of both the reference and impaired point clouds.

Finally, the three similarity measures, i.e., SIM_{m_g}, SIM_{μ_g}, and SIM_{c_g}, are obtained as follows:

$$SIM_{m_g} = \frac{2m_g^r \cdot m_g^d + T_0}{(m_g^r)^2 + (m_g^d)^2 + T_0} \quad (2.16)$$

$$SIM_{\mu_g} = \frac{2\mu_g^r \cdot \mu_g^d + T_1}{(\mu_g^r)^2 + (\mu_g^d)^2 + T_1} \quad (2.17)$$

$$SIM_{c_g} = \frac{c_g d + T_2}{\sigma_g^r \cdot \sigma_g^d + T_2} \quad (2.18)$$

where T_0, T_1 and T_2 are non-zero constants defined to prevent numerical instability, set to 0.001. The overall similarity is given by $S_{\overrightarrow{s}_k,C} = SIM_{m_g} \cdot SIM_{\mu_g} \cdot SIM_{c_g}$, which is then pooled across all color channels:

$$S_{\overrightarrow{s}_k} = \frac{1}{\gamma} \sum_C \gamma C \cdot |S_{\overrightarrow{s}_k,C}| \quad (2.19)$$

In equation (2.19), γC is the pooling factor that reflects the importance of each color channel in the visual perception. The point cloud RGB components are decomposed to the Color Gaussian model [37]. This results in a luminance component ($\hat{E}$) and two chrominance components ($\hat{E}_\lambda$ and $\hat{E}_{\lambda\lambda}$). As such, the authors set the pooling factor as $\hat{E} : \hat{E}_\lambda : \hat{E}_{\lambda\lambda} = 6 : 1 : 1$, as in the overall PSNR calculation of YUV [95]. The final overall similarity score is obtained by averaging across the total number of keypoints. Finally, for

each scale, the overall quality scores are aggregated using the following pooling operation:

$$Q_{overall} = \frac{\sum_{i=0}^{M} w_i S_{\vec{s}_k}}{\sum_{i=0}^{M} w_i} \quad (2.20)$$

where M represents the different scales and w_i denotes the weighted factor for the different scales.

2.4.1.7 Point to Distribution [98]

This metric uses the *Mahalanobis* distance in order to measure the distance between a point and a distribution of points [99, 98]. This metric makes use of the scale-invariant property of the *Mahalanobis* distance in order to assess the geometry and color point to distribution distortions. Then those distortions are fused, obtaining a joint geometry and color quality metric. After defining a neighborhood for each reference point in the distorted point cloud the Mahalanobis distances are computed for the geometry and color components.

These *Mahalanobis* distances are defined in a neighborhood and are characterized by the average and standard deviation to the reference point. Then, for those distances an average pooling for each reference point is obtained. The process is repeated by swapping the reference and distorted point clouds. The maximum pooling distance from each process is selected.

This is computed for the geometry and for the color components described in YUV. For the three color components the *Mahalanobis* distance are computed. The Point-to-Distribution (P2D) results can be computed using the weighted P2D of the three components. If all three components are considered, the Y,U and V have weights of 6, 1 and 1 respectively as it is common. Alternatively, only the Y color channel can be used.

Finally the P2D resulting from the geometry and from the color are averaged resulting in the final metric value.

2.4.1.8 Color Histogram Metric [100]

The metric extracts color features from reference and distorted point clouds. As color histograms represent the probability distribution of pixel values for the entire volume, distortions applied to the color channel usually modify the distribution of colors. The color histogram is computed for both the reference and the distorted point clouds, and the histogram distance is measured. A weighted average for the YCbCr is also applied, as shown in equation 2.21:

$$Histogram = \frac{6 \times dist_Y + dist_{C_b} + dist_{C_r}}{8} \quad (2.21)$$

2.4.2 Reduced Reference Metrics

2.4.2.1 Reduced Reference Point Cloud Metric (PCM_{RR}) [21]

This reduced-reference metric extracts a small set of geometry and attribute related features, which are used to predict the visual quality of the content under evaluation, finding the best combination of features [21]. For geometry features, a set of coordinates is defined, for the x, y, z axis. For each axis, a vector containing features is computed, obtained through max pooling. This results in 7 features, which are the mean, standard deviation, median, mode, entropy, energy and sparsity. The attribute related features are computed in the luminance channel. The color space is converted from RGB to YUV, and then a features set is extracted, defined by the same ones described in the geometry features. This metric also considers normal based features. For each point with a normal attribute, the $K-$nearest neighbors are selected, and then the angular similarity between those points is obtained. This results in a matrix with two dimensions, so the feature vector is redefined, having 7 features as well, which are mean of means, mean of standard deviation, mean of medians, standard deviation of means, entropy, energy sparsity. Defining the three sets of features as Φ^G for geometry features, Φ^L for luminance features and Φ^N for normal features, extracted from a distorted point cloud, they are compared to the features extracted from the reference point cloud. For each pair of features, with $\hat{f}_i$ describing the distorted cloud features and f_i the original cloud features, the absolute difference d_i is calculated, and the final score is obtained as:

$$\text{PCM}_{\text{RR}\cdot} = \sum_i w_i d_i \quad (2.22)$$

where $w_i \in [0, 1]$ are the weights, obtained and validated via training on a point cloud set. The training was made by the authors using the dataset available in [80]. The previously described features were extracted, and a linear optimization algorithm was used to maximize the Pearson Linear Correlation Coefficient between the metric and the scores, after logistic fitting.

2.4.2.2 Reduced Reference Content-oriented sAliency Projection - RR_{CAP}

To reduce the reference point cloud, the downsampled saliency maps are extracted, after view projection. To project the views to a 2D plane, six perpendicular projections are adopted. After projecting the views the saliency maps are extracted by image signature [101]. The image is downsampled to a coarser counterpart. Then, a sign function of discrete cosine transform coefficients is specified as image signature, for the downsampled projection. Afterwards, the image signature, now in transformed format, can be converted back to the spatial domain by inverse DCT, and the salience map is computed

as:

$$m = IDCT() \odot IDCT() \quad (2.23)$$

where $\odot$ indicates the Hadamard product, and represent the image signature for the downsampled projection.

The image structural similarity between the reference and test projected saliency maps. The structural similarity between the saliency maps obtained for the reference and distorted point clouds is computed by equation 2.24:

$$S = \frac{(2\mu_r\mu_d + C_1)(2\sigma_{rd} + C_2)}{(\mu_r^2 + \mu_d^2 + C1)(\sigma_r^2 + \sigma_d^2 + C2)} \quad (2.24)$$

where the $\mu_r, \mu_d, \sigma_r^2, \sigma_d^2$ and σ_{rd} represent the local mean, variance and covariance, of the reference and distorted saliency maps. Additionally, C_1 and C_2 are stabilizing constants.

To complement pooling similarities, a content-oriented weighting strategy is employed. To perform this, it is considered that the content is quantified by spatial information, using the Sobel filter. Spatial information is used as the weight for the content-oriented similarity and computed by equation 2.25:

$$w = |sd[Sobel(I_d)] - sd[Sobel(I_r)]| \quad (2.25)$$

where $sd[\cdot]$ is the standard deviation operated over the image pixels, and I_r and I_d are the reference and distorted projected imaged, respectively. The content oriented similarity is then obtained by equation 2.26:

$$S_w = \frac{1}{n}\Sigma S^w \quad (2.26)$$

whew n is the number of projections.

This metric also used statistical information collected from the saliency maps. For both the reference and distorted point cloud, the statistical correlation given by equation 2.27:

$$H_c = \frac{1}{n}\Sigma \frac{E[h_r h_d] - E[h_r]E[h_d]}{\sqrt{(E[h_r^2] - E[h_r])^2}\sqrt{(E[h_d^2] - E[h_d])^2}} \quad (2.27)$$

where $E[\cdot]$ represents the expectation operator, and h_r and h_d the statistical histograms of the reference and distorted saliency maps. The two quality measurements S_w and H_c are combined to an objective quality score using equation 2.28

$$Q = S_w \cdot H_c \quad (2.28)$$

2.4.3 No-Reference Metrics

2.4.3.1 ResSCNN [91]

This metric is composed by three modules, namely a hierarchical features extraction module W^f, a pooling and concatenation module Φ and a quality prediction module M^r. The module W^f takes the reference point cloud and extracts hierarchical features using a stack of sparse convolutional layers and residual blocks. The Φ module generates feature vectors by conducting global pooling and concatenation operations. Finally, W^r predicts the quality scores using the features vectors computed by Φ [91].

The hierarchical extraction of features consists of four different blocks, each containing three sparse convolution layers. The second and third layers are connected to the residual pattern. The output features computed by the w^f module may have different shapes. To overcome this, the extracted features from the sparse convolutional neural network (CNN) are globally pooled into a 64×1 feature vectors. They are then concatenated to generate the 256×1 representative hierarchical feature vector. The global pooling is then applied, obtaining the feature vector from different depths of the sparse CNN, using equation 2.29:

$$S_i = \Phi_i(X; w^f), i)1, ..., 4 \quad (2.29)$$

in which $S_i \in R^{64\times1}$ represents the normalized feature vector, computed in the various depths of the hierarchical feature extraction module, W^f represents the several parameters of the hierarchical feature extraction module, and finally Φ_i is the global pooling operations computed on the features with irregular shapes. After pooling, the normalized S_i vectors are concatenated to a vector S, and the final predicted quality scores as computed using equation 2.30:

$$Q = F(S; W^r) \quad (2.30)$$

where Q is the final quality score, F represents the fully connected layers and W^r the parameters of the quality prediction model [91].

The loss function smooth L_1 is adopted, and is depicted in equation 2.31.

$$SmoothL_1(x) = \begin{cases} 0.5x^2, & if|x| < 1, \\ \pm1, & otherwise \end{cases} \quad (2.31)$$

In the equation, $x = q - \hat{q}$, where q is the predicted quality score by the model, and $\hat{q}$ is the ground truth. To accelerate the speed of training, the Stochastic Gradient Descent (SGD) is employed, with a learning rate of $1e - 3$.

2.4.3.2 Image Transferred Point Cloud Quality Assessment (IT-PCQA) [102]

IT-PCQA consists of a point cloud pre-processing module, the feature generative network G, the conditional discriminative network D, and the quality regression module R.

The pre-processing module aims to transfer point clouds to multi representative images. To achieve this, the point clouds are projected onto six planes of a cube, and the images are spliced to form a multi-perspective image. This image will be used as input to the network G.

The network is comprised of a SCNN network as a feature encoder, adapted with a hierarchical component, as the SCNN is not very computationally expensive [103]. Nine convolutional layers are introduced, followed by batch normalization and rectified linear unit (ReLU) operations, resulting in a sketching feature map. To generate the final feature map, the 3rd, 5th, 7th and 9th layered features and extracted. After average pooling, four latent features are obtained and concatenated. Finally, a twofold convolution is conducted to generate the final features. This process is shown by equation 2.32:

$$G(x) = \Gamma\{\Phi_3(x) \oplus \Phi_5(x) \oplus \Phi_7(x) \oplus \Phi_9(x)\} \quad (2.32)$$

where Φ_j represents average pooling, $\oplus$ represents concatenation and $\Gamma\cdot$ represents the final twofold convolutions.

The resulting features are then fed to a conditional-discriminative Network, which identifies features that are not fit to predict quality scores, using a method based on SROCC [102]. The quality regression network uses twofold full connection layers in order to regress an objective score from a latent feature generated by both G and D.

2.4.3.3 Multi-Modal Point Cloud Quality Assessment (MM-PCQA) [104]

In an initial step, this metrics segments the input point clouds to sub models. To conduct this, the point clouds are normalized, resulting in the normalized point cloud $\hat{P}$. The farthest point sample methodology is used to obtain N_δ anchor points $\{\delta\}_m^{N_\delta} = 1$. For each of those anchor points, the K nearest neighbor (KNN) algorithm is employed to generate N_s neighborhoods around N_δ. The neighborhoods are converted to sub-models using equation 2.33

$$S = KNN(\delta_m)_{m=1}^{N_\delta} \quad (2.33)$$

where S is the set of sub-models and δ_m the m-th farthest sub-sampling point.

A point cloud feature encoder is $\theta_p(\cdot)$ is implemented, that maps the obtained sub-models to a quality-aware embedding space [104].

In parallel, N_I image projections are rendered from the point cloud, using a defined circular pathway with fixed viewing distance. The projections are rendered to the quality-aware [104] space with a 2D image encoder θ_i.

A symmetric Cross-Modality Attention block is employed to assess the interaction between the visual quality features gathered from both modalities (point cloud and 2D images). To employ this block, the features generated by θ_p and θ_i are adjusted to the same dimension, using linear projection. The final quality feature, results from the concatenation of the intra-modal features, and the obtained by the cross-modality attention block. To optimize the quality regression model, the MSE is employed [104].

2.4.3.4 NR-3DQA [105]

This metric is a no-reference metric that employs machine-learning, namely a Support Vector Regression model [106], trained with several geometry and color features extracted from the point cloud. Those features are based on curvature, anisotropy, linearity, planarity and sphericity [105]. Overall seventy seven features are extracted, and then fed to a SVR, in order to train it, providing a quality score.

2.4.3.5 Video Quality Assessment metric (VQA) [107]

VQA aims to assess the quality of point cloud distortions by projecting the input point cloud to a video. The camera is placed at a default position with a fixed viewing distance. Then, four distinct pathways are defined that defined symmetric circular rotations, namely a rotation on the XY and YZ plans, as well a tilt for 45 deg relative to both the XY and XZ planes. To generate consistent videos, the camera rotation step is set to 12 deg between consecutive frames. 30 frames are extracted for each circular pathway, resulting in a 120 frames videos.

After generating the videos, the quality-aware features from the temporal and spatial domain are extracted. From among the j frames in a given i video, a key frame K_i^j is firstly selected using the viewpoint-max-distance method. It selects the frames with the farthest viewpoints between each other. A 2D-CNN model is employed to extract the spatial feature maps.

The temporal features are more focused in the relation between continuous frames, and are considered to be more capable of reflecting perceptual distortions [107]. To do this, a pre-trained 3D-CNN model is employed to extract the temporal features. Before this procedure, the videos are downsampled to a low resolution.

The features are then merged using a concatenation operation. A two stage full-connected layers is employed, with the first layer containing 128 neurons and the second layer containing 1. The final quality score is represented by equation 2.34:

$$Q = \frac{1}{i}\Sigma_{i=1}^{n} Q_i \quad (2.34)$$

where Q is the predicted quality for the video, and i is the number of videos. Furthermore, the metric uses the MSE as loss function.

2.4.4 Image quality metrics

Traditional image quality metrics, such as the PSNR, SSIM [108], MS-SSIM [109], VIFp [110], FSIM [111] and FSMc [111], VMAF [112] or LPIPS [113] can be employed for objective quality assessment of point clouds [18, 5, 94]. If 2D the projections of the point cloud are extracted, the above metrics can be computed. However, these metrics depend on the visualization directions of the point clouds, leading to some instability. Furthermore, some recent works [18, 5, 94] seem to indicate that the most recent point cloud metrics tend to provide better performance.

2.5 Benchmarking objective quality metrics

To benchmark objective quality metrics, the scores are normalized between 0 and 1. Then, the normalized scores, as well as the normalized MOS results, are used to fit a logistic function. The fitted function is then used to predict MOS values.

The fitted function (shown in equation 2.35) is the most commonly used to predict MOS values.

$$f(x) = a + \frac{b}{1 + \exp^{-c\cdot(x-d)}} \quad (2.35)$$

The predicted values are compared to the MOS values, obtained in the subjective quality evaluation, using the statistical indicators specified in ITU-T P.1401 [26], as is commonly done when benchmarking objective metrics [114, 115, 5].

The specified statistical indicators are the Pearson Correlation Coefficient (PCC), Spearman Rank Order Correlation Coefficient (SROCC), Root Mean Squared Error (RMSE) and the Outlier Ratio (OR).

PCC [116, 117] measures the linear correlation between two sets of data, and it is used to assess the degree of agreement between the subjective scores and predicted scores by a certain objective quality metric.

SROCC [118] measures the monotonic relationship between two variables by comparing the ranks of the data. It is important to understand how the order of quality levels is

preserved between subjective and objective scores.

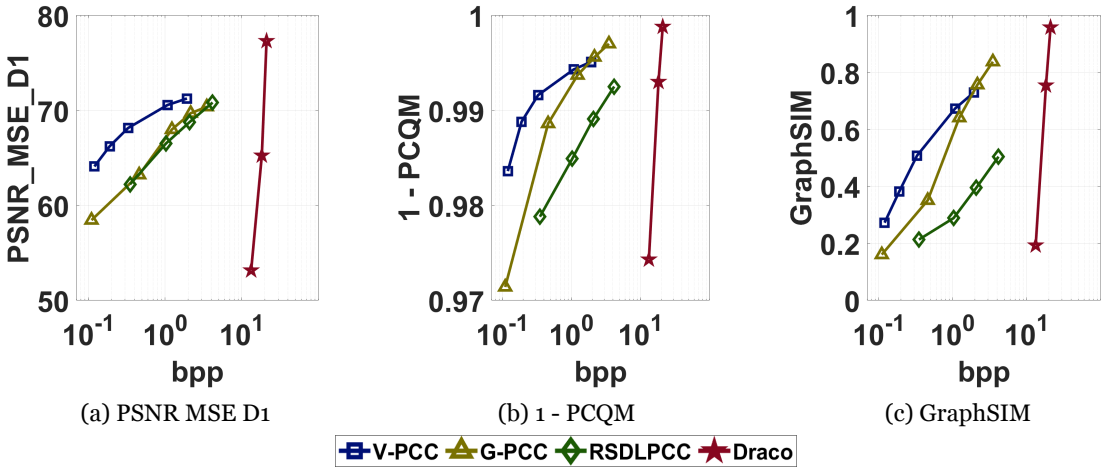
RMSE [119] measures the average magnitude of the error between the subjective and objective scores, providing a direct indication of how closely the objective scores approximate the subjective scores.

Finally, OR [120] computes the proportion of scores that deviate significantly from the expected values. It aids in detecting and quantifying the presence of anomalous ratings that differ from the majority of subjective scores.

Although Kendall's Tau [121] is a valuable statistical tool for measuring rank correlation, it is not present in the ITU-T P.1401, and it is not commonly used in most subjective quality evaluation studies. This can be attributed to the redundancy with SROCC.

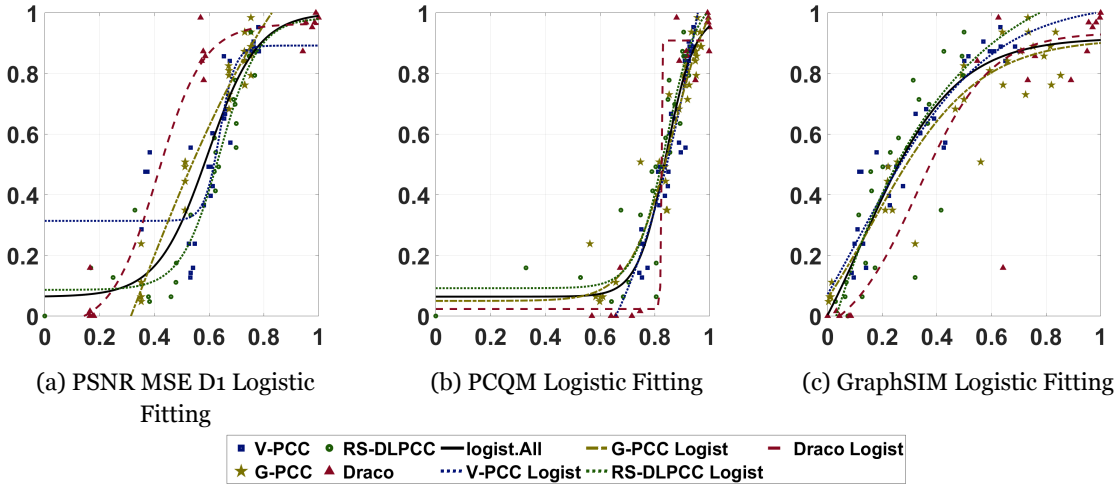
Fig. 2.19 shows three examples of objective metrics obtained for the *Longdress* point cloud that has the subjective results shown in Fig. 1.8. It is possible to observe that *1-PCQM* and *GraphSIM* show a very similar behavior with the MOS, while PSNR MSE D1 does not reflect the MOS in the same manner.

Fig. 2.20 shows the logistic fitting plots obtained with eq. 2.35 and the location of the pairs normalized MOS (y axis) versus metric values (x axis). It is possible to observe that the points representing the MOS and predicted MOS are more concentrated near the regression curve for PCQM. For the PSNR MSE D1, the pairs MOS/predicted MOS are more disperse. The GraphSIM exhibits an intermediate behavior that lies between the two.



(a) PSNR MSE D1 (b) 1 - PCQM (c) GraphSIM

Figure 2.19: Metric vs bpp plots for the *Longdress* point cloud (Plots taken from [5]). The subjective evaluation results are represented in Fig. 1.8.



(a) PSNR MSE D1 Logistic Fitting (b) PCQM Logistic Fitting (c) GraphSIM Logistic Fitting

Figure 2.20: Example of objective quality metrics fitting with eq. 2.35. the y axis represents the normalized MOS and the x axis represents the respective metric values.

Chapter 3

Methodologies for quality evaluation

This chapter describes the methodologies that were considered in the doctoral program. It contains information on the steps to consider when conducting subjective evaluation, as well a description of the most commonly used objective quality metrics. Furthermore, visual examples are included, and insights on some of the conducted studies, is also provided.

3.1 Subjective evaluation

This section describes the usual steps taken when preparing a subjective quality evaluation study.

3.1.1 Selection of a dataset

Selecting the dataset for evaluating point cloud coding solutions is of the utmost importance. A rich dataset helps to understand how the codecs behave under different types of point clouds. as well as dense and solid point clouds. Across the conducted experiments, point clouds with different densities were considered. Very sparse point clouds, such as the ones obtained by LiDAR technology, were not considered, as they are not fit for subjective quality evaluation. Furthermore, it is important to consider point clouds with a wide range of colors, to evaluate how efficient and reliable is the coding of the color information.

Several different characteristics are considered when selecting point clouds for subjective evaluation.

- Sparsity: Defined as the average distance between each point and its 20 nearest neighbors.
- Color Gamut Volume: The volume of the Convex Hull of the distribution of color points in CIE 1976 LAB space divided by the volume of the CIE 1976 LAB color space[1]
- Curvature Statistics: For each point, local curvature is computed using the method described in Taylor *et al.* [122].

[1] http://www.brucelindbloom.com/index.html?WorkingSpaceInfo.html

Table 3.1: Characteristics of the JPEG Pleno Point Cloud test set [17].

Name	Number of points	Geometry Precision	Density Class	Density Factor	Homogeneity Class	Homogeneity Factor	Color Gamut Volume
Thaidancer	3130215	12	solid	3.28E-01	homogeneous	5.000	22.29%
soldier	1089091	10	solid	4.18E-01	homogeneous	8.333	1.18%
RWT70-StMichael	1871158	10	solid	4.18E-01	homogeneous	6.897	21.22%
RWT130-Bouquet	3150249	10	solid	4.18E-01	homogeneous	10.345	41.05%
Facade2	1596085	12	dense	1.39E-02	heterogeneous	13.208	5.03%
House_without_roof	4848745	12	dense	3.65E-02	heterogeneous	37.415	12.93%
boxer	3493085	12	dense	4.85E-02	homogeneous	11.111	2.63%
CITIUSP	5705126	13	dense	2.65E-02	heterogeneous	22.642	29.27%
Arco_Valentino_Dense	1481746	12	sparse	4.13E-03	homogeneous	7.066	14.83%
Shiva	1009132	12	sparse	3.77E-03	homogeneous	11.304	19.37%
ULB_Unicorn	1995189	13	sparse	5.22E-04	heterogeneous	12.500	50.50%
EPFL	4694733	13	sparse	9.95E-03	heterogeneous	17.390	18.52%

- Local Density (LD): is computed by counting the number of neighbors (N) inside a sphere with radius R, for each point. Then, N is divided by the neighborhood volume. This is defined in the following equation:

$$DL = \frac{N}{\frac{4}{3} \times \pi \times R^3} \quad (3.1)$$

- Global Density: A histogram can be computed from the local densities of each point, as well as the median and the interquartile range (IQR). The median can be used as an estimate for the global density DGM.

- Spread and homogeneity: The local density can also be used to determine the degree of homogeneity of a point cloud. Homogeneous point clouds present a low local density variation, while heterogeneous point clouds present high local variation.

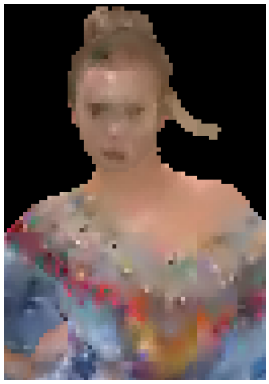
Table 3.1 describes the characteristics of the JPEG Pleno Point Cloud test set. The test set comprises 12 point clouds with bit depth varying from 10 to 13. Furthermore, the set is divided in 4 solid, dense and sparse point clouds. Finally, the Color Gamut Volume ranges from 1.18% to 50.50%.

3.1.2 Dataset Coding

To prepare a subjective evaluation, the point clouds in the selected dataset are coded with the codecs under evaluation. Usually, four or five distinct quality levels are chosen. In the JPEG Pleno Learning Based Point Cloud Coding CTTC document [17], are defined

Table 3.2: Target bitrates according to the JPEG Pleno Learning Based Point Cloud Coding CTTC document [17].

	R01	R02	R03	R04
geometry only	0.05	0.15	0.5	1.5
geometry + color	0.1	0.3	1	3



(a) G-PCC (R01)

(b) G-PCC (R03)

(c) G-PCC (R05)

(d) V-PCC (R01)

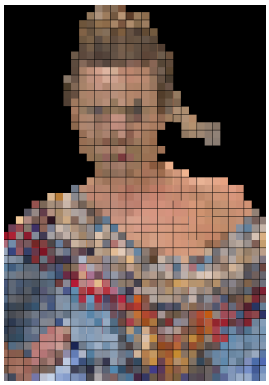
(e) V-PCC (R03)

(f) V-PCC (R05)

(g) RS-DLPCC (R02)

(h) RS-DLPCC (R03)

(i) RS-DLPCC (R05)



(j) Draco R01

(k) Draco R03

(l) Draco R05

Figure 3.1: Crop area of decoded results for *Longdress* (Images taken from [5])

the target bitrates of Table 3.2 for geometry coding only and for both geometry and color coding. It is expected that these bit rates will result in a diversity of coding distortions that result in perceptual qualities ranging from very low to very high quality, both for the anchors, and for the call for proposal responses. Sometimes, it is not possible to achieve those desired bitrates, since some codecs do not allow an easy bit rate control.

The selected bitrates should be somehow equidistant in a logarithmic scale, which usually results in a linear quality evolution. This usually ensures that the changes in quality are visually perceptible. Fig. 3.1 shows examples of a point cloud (*Longdress)* coded with four different codecs at three different quality levels. The typical distortions caused by the codecs can be observed, such as the reduction in resolution typical of octree-based codecs (G-PCC or Draco, that is based in KD-tree), the typical smoothing that can be found on projection-based codecs (V-PCC), and the appearance of empty spaces typical of machine-learning-based codecs (RS-DLPCC).

3.1.3 Visualization of point cloud content

Unlike 2D images, point clouds are represented by spheres centered in each point coordinates. This representation can lead to empty spaces that allow the visualization of the inner part of the point cloud, resulting in a negative impact on the subjective scores [75, 78]. For an appropriate visualization of the point cloud content, the sphere diameter that represents each point needs to be increased to create the sense of a continuous surface. The diameter of the sphere is in the following defined as point size. This manipulation is critical to avoid the perceptual effect described above. However, it must be ensured that this manipulation does not mask compression artifacts. An example is represented in Fig. 3.2, namely for the *Longdress* point cloud, encoded using G-PCC. It can be observed that without manipulating the point size (Fig. 3.2b), parts of the point cloud are missing, and the opposite part of the point cloud is visible. This tends to happen when encoding at low rates, as the points are usually rather far from each other. When the point size is increased (Fig. 3.2c), the artifacts created by the codec are still noticeable, but the inner part of the point cloud is not visible anymore.

3.1.4 Stimuli Generation

Typically, subjective quality evaluation is conducted using 2D displays [81, 16] in a controlled environment. Besides a 2D display, stereoscopic 3D displays [14], as well as a Head Mounted Display (HMD) can be used for subjective quality assessment.

(a) Reference: Point size 4

(b) G-PCC (0.11 bpp): Point Size 4

(c) G-PCC (0.11 bpp): Point Size 12

Figure 3.2: Point size example for the *Longdress* point cloud.

3.1.4.1 Using a 2D display

For the subjective quality evaluation in 2D displays the point cloud needs to be show in different viewing positions. For that a video is created where the reference and distorted point clouds are rotated around a selected axis in a video side by side for a double stimulus evaluation. For all point clouds, a complete rotation over the vertical axis was applied. For each rotation degree, an image representing the projection of a point cloud view is extracted. Those views can be obtained using the Point Cloud Library (PCL)[2] Visualizer mode, or Cloud Compare[3]. The videos are created using the FFMPEG[4] software. A lossless compression using H.264 [123] is employed, to ensure that no video compression artifact is added to the point cloud coding. To ensure that lossless coding is applied to the extracted frames, the Constant Rate Factor (CRF) is set to 0. Alternatively, the copy parameter can also be used. Furthermore, it should be ensured that there is no conversion to the YUV color space. This is conducted by using the `libx264rgb` flag in FFMPEG. The point cloud views are rendered at 30 fps. Each view represented a 1° rotation, resulting in 12 second videos displayed with a resolution that should be the same as the display that is being used. The typical command to generate videos with ffmpeg is as follows:
`ffmpeg -framerate 30 -start_number 0 -i inputFrames -c:v libx264rgb -crf 0 outputVideo.mp4`.
Fig. 3.3 shows an example of a frame taken from a video used in the study conducted by Prazeres *et al.* [13]. The videos are shown directly in a 2D display, without any kind of post processing, avoiding the addition of any new artifact or the masking of artifacts caused by the point cloud compression.

[2]https://pointclouds.org
[3]https://www.danielgm.net/cc/
[4]https://ffmpeg.org/ffmpeg.html

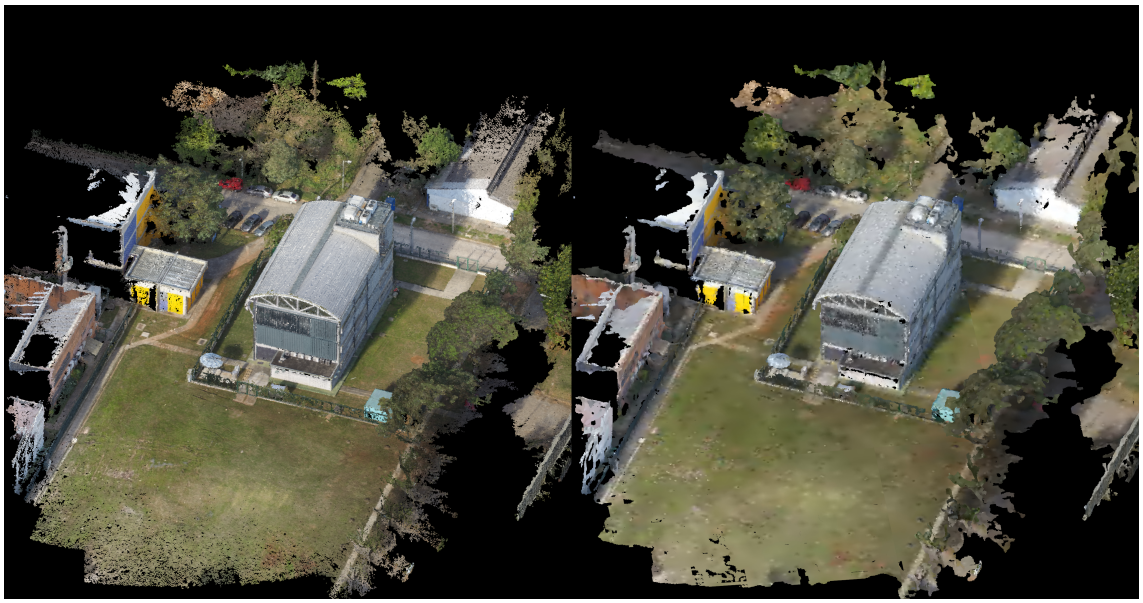


Figure 3.3: Example frame from a 2D video used in 2D subjective evaluation [13]. In this case, the reference point cloud is on the left side, and the distorted point cloud is on the right side.

3.1.4.2 Using a 3D stereoscopic display

When conducting subjective quality evaluations using a 3D stereoscopic display, two different projections need to be created for each eye. These two projections are required for both the reference and distorted point clouds. Firstly, the point clouds were scaled in the X and Z axes using the PCL library. The left and the right representations are projections of the point cloud shifted by $1.5°$ and $-1.5°$ considering as reference the frontal view (as used in [16]).

The final frames are composed of the left view distorted, left view reference, and right view distorted, right view reference, for visualization of the 3D with the distorted point cloud on the left side. For the visualization with the reference on the left side, the reference and distorted were swapped. This is needed because two subjects doing tests in a row will show the reference and the distorted swapped reference (reference appears once on the left and once on the right side of the screen). To create the proper stereoscopic effect, a translation also needs to be applied to each content, given by equation 3.2

$$N_x = \frac{\frac{S_x}{2}}{\Delta}, \quad \text{with} \quad S_x = Z_{rec} \times \frac{IPD}{Z_v} \tag{3.2}$$

where Z_{rec} is the camera captured distance of the point cloud, Z_v is the visualization distance of the point cloud, IP is the interpupillary distance, and Δ is the horizontal length corresponding to a sample in the camera. The 3D volumes were located near the display plane to reduce the fatigue caused by the stereoscopic visualization. Fig. 3.4 shows a frame taken from a video used in the subjective evaluation conducted by Prazeres *et al.* [36].

Figure 3.4: Example frame from a 3D video used in 3D stereoscopic subjective evaluation [14]. In this case, the reference point cloud is on the left side, and the distorted point cloud is on the right side.

3.1.4.3 Using a Head Mounted Display (HMD)

To visualize point cloud content using an HMD, the Unity[5] software is used [85]. The Pcx point cloud importer library[6] allows the manipulation and visualization of point cloud data. The point clouds are positioned at a distance that does not cause discomfort to the subjects, that were seated in a fixed position. Fig. 3.5 shows a point cloud rendered in unity, using the Pcx package. One advantage of using unity is that it is not necessary to generate the visualization path as a video content. Because of that the complexity and resources required for the testing are largely reduced since the typical videos used in the previously described subjective quality evaluation modalities are uncompressed, and have a 4K resolution in most of the cases.

3.1.5 Texture mapping

To conduct subjective quality assessment of coding solutions that code only geometry information, it is necessary to, in some way, add texture information to the coded point clouds, as it is very important for the quality perception. During the doctoral program, two different experiments were conducted, namely Evaluation 1 and Evaluations 2, with two different methods of adding texture information to the coded point clouds [124]. In both experiments, the ADLPCC [53], PCC GEO CNNv2 [7], PCGCv2 [6] and LUT SR [34]

[5]https://unity.com
[6]https://github.com/keijiro/Pcx

Figure 3.5: Point cloud in Unity

were selected. The codecs were used to code six distinct point clouds, targeting five different quality levels, ranging from poor to high quality.

In *Evaluation 1*, the texture information of the reference point cloud was mapped to the distorted geometry. Then, the resulting point cloud (containing the distorted geometry

Table 3.3: Correlation of the objective metrics with the subjective quality evaluation results. The best values are shown in bold, and the second best values are shown in italic.

		Evaluation 1			
Metric	Type	PCC	SROCC	RMSE	OR
MSE PSNR D1	FR, GEO	0.806	0.782	0.184	0.753
MSE PSNR D2	FR, GEO	0.821	0.796	0.177	0.813
PointSSIM	FR, COL	*0.859*	*0.857*	*0.159*	*0.720*
Point 2 Distribution	FR, $GEO + COL$	0.851	0.828	0.164	0.640
PCM-RR	FR, $GEO + COL$	0.834	0.834	0.172	0.727
GraphSIM	FR, $GEO + COL$	0.800	0.799	0.186	0.780
PCQM	FR, $GEO + COL$	**0.899**	**0.903**	**0.137**	**0.573**
		Evaluation 2			
MSE PSNR D1	FR, GEO	**0.834**	**0.774**	**0.152**	**0.720**
MSE PSNR D2	FR, GEO	*0.777*	*0.740*	*0.174*	*0.793*
PointSSIM	FR, COL	0.188	0.143	0.271	0.920
Point 2 Distribution	FR, $GEO + COL$	0.437	0.472	0.249	0.873
PCM-RR	FR, $GEO + COL$	0.408	0.323	0.252	0.900
GraphSIM	FR, $GEO + COL$	0.560	0.573	0.229	0.907
PCQM	FR, $GEO + COL$	0.634	0.700	0.214	0.787

and the texture information of the reference point cloud), was encoded with G-PCC using the `lossless-geometry-lossy-atts` mode. This ensures that the no further artifacts are introduced by G-PCC in the distorted geometry.

In the second subjective quality evaluation (*Evaluation 2*), the data preparation was similar, but instead of encoding texture using G-PCC, the texture of the reference point clouds was mapped directly onto the decoded geometry.

For both experiments, the texture was mapped using *Meshlab*[7]. The software maps the color using the nearest neighbor algorithm. For the point cloud without texture information, the nearest neighbor of the point cloud with texture information is identified. The color of the nearest neighbor is then assigned to that point.

Table 3.3 summarizes the corresponding correlation measures for both experiments. In the table, the best values are shown in bold, and the second-best values are shown in italic.

The study concluded that encoding the texture (Evaluation 1) provides a more reliable subjective evaluation than just mapping the reference texture to the resulting geometry. Furthermore, metrics reveal a typical performance when the texture is encoded (Evaluation 1), while their performance is unstable and unusual when texture is directly mapped onto the decoded geometry (Evaluation 2).

3.2 Benchmarking Objective Quality Features

Objective point cloud quality metrics are usually defined as a set of features that are then somehow combined to provide a final quality score [15, 125].

3.2.1 Feature Analysis

The importance each feature has to the final quality estimation can be analyzed using an algorithm such as the Recursive Feature Extraction (RFE) algorithm. The most important features can then be combined using regression algorithms such as Support Vector Regression (SVR) [106], Ridge Regression (RiR) [126] or AdaBoost [127].

Fig. 3.6 shows the framework of this methodology. The features of the five different objective quality metrics, namely PSNR MSE D2, PCQM, MS-GraphSIM, PointSSIM and PSNR MSE YUV are extracted from both the reference and distorted point clouds, obtaining a combined feature vector. That vector will be analyzed using RFE with a regression model to obtain a ranking of the most important features. Finally, the most important features are selected and used as input to the regression model used to rank the features, leading to a final quality estimation score.

[7] https://www.meshlab.net

An initial study was conducted in order to assess the importance of several features defined in point cloud quality metrics [15] . The study employed RFE to rank the features in order of importance, and the SVR and RiR regression algorithms were selected.

SVR [106] aims to find a hyperplane that fits better the data while allowing for some margin of error.

RiR[126] adds a penalty to the size of coefficients in a linear model to prevent overfitting by discouraging large coefficients.

The results of feature ranking for both algorithms is shown in table 3.4. It reveals that features that consider luminance information are the most prominent ones.

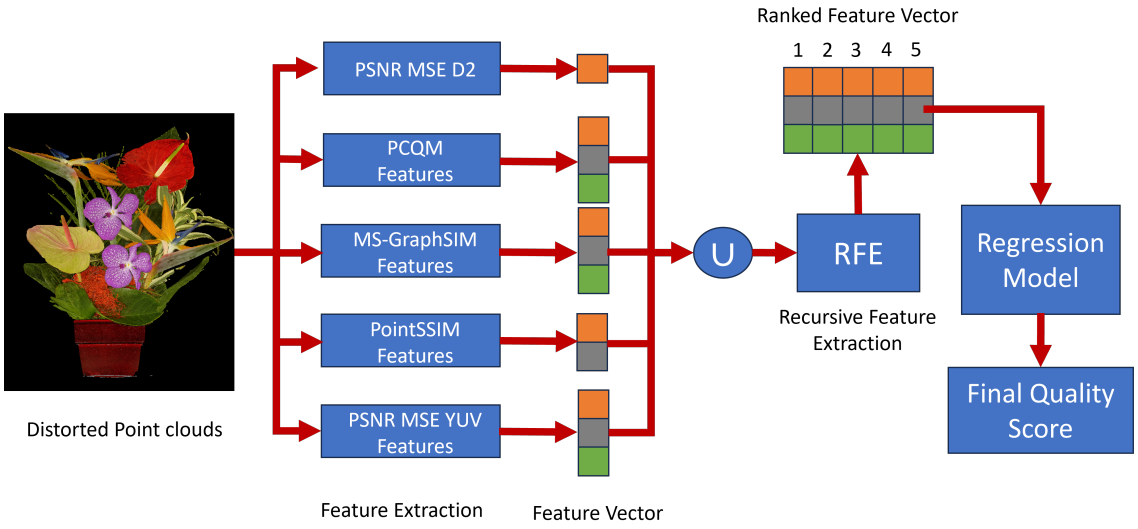


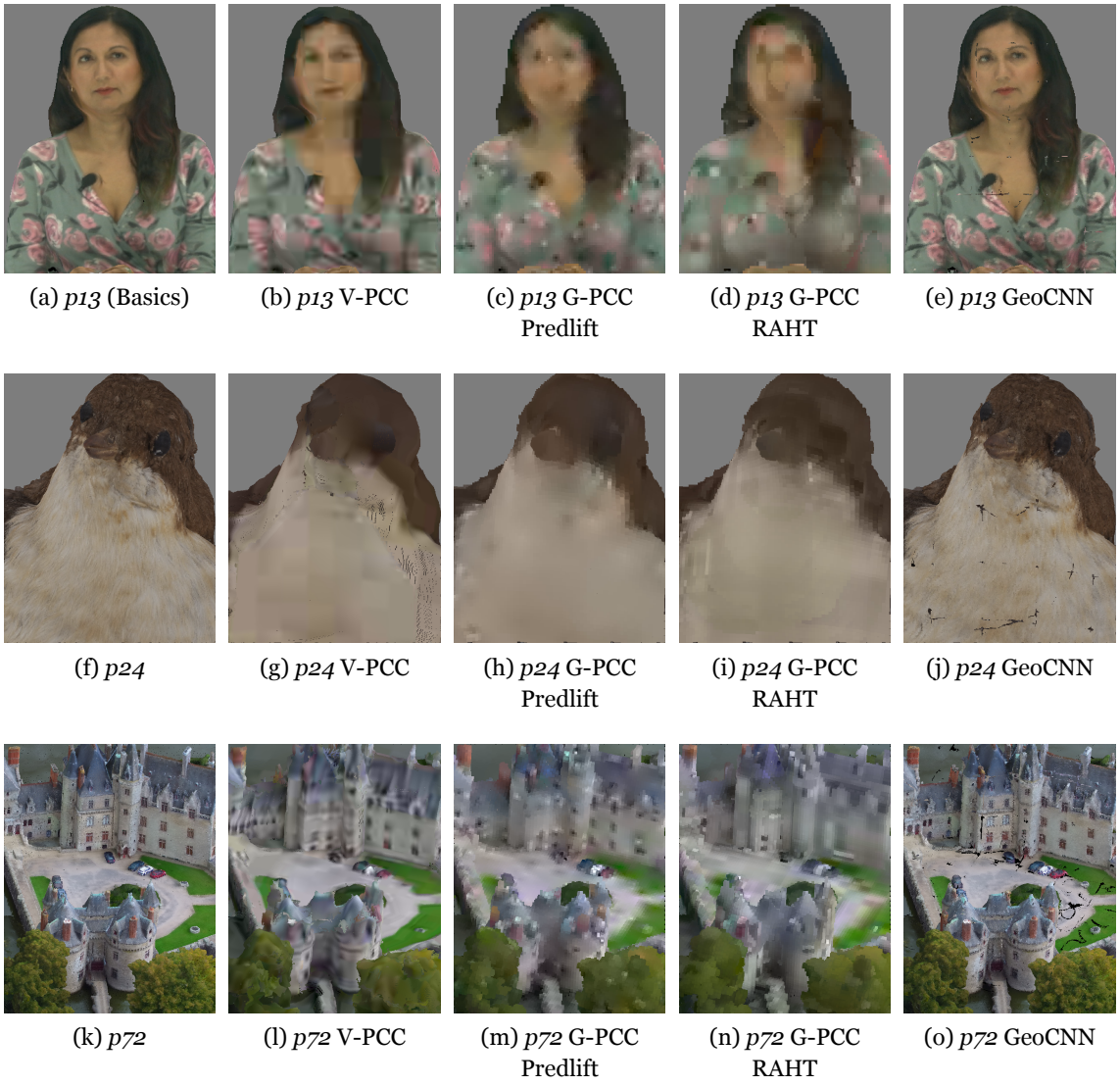
Figure 3.6: Framework of the feature extraction study and regression model. First, the features of the considered metrics are extracted, obtaining a vector with all the computed features. The importance of each feature is then analyzed using RFE, resulting in a Ranked Feature Vector. Finally, the quality scores are given by a selected regression model.

3.2.2 Dataset for feature study

The Broad Quality Assessment of Static Point Clouds in Compression Scenario (BASICS) training dataset [18] was selected to study the contribution of each feature for the prediction of objective quality scores. This database contains 898 coded point clouds, with distortions introduced by different coding methods, notably the octree model of G-PCC [1], using both the RAHT [49] and Predlift [1] methods, the video-based codec V-PCC [1], and the learning-based solution GeoCNN [7], from 45 reference point clouds. Furthermore, the point clouds represent several scenarios. Fig. 3.7 shows three examples of point clouds from BASICS, representing human content (*p13*), a bird (*p24*), and a landscape (*p72*). The results obtained for the lowest coding rates are shown, so that the distortions typically created by these aforementioned coding solutions are well visible. This dataset was chosen

Table 3.4: Feature Ranking using RFE for the BASICS Database [18].

	Feature	
Ranking	SVR	RiR
1	PSNR MSE D2	
2	MS-GraphSIM SIM_{m_g} Scale 0	
3	PCQM $f4$	
4	PCQM $f5$	
5	MS-GraphSIM m_g Scale 2	PCQM $f2$
6	PCQM $f2$	PCQM $f7$
7	MS-GraphSIM SIM_{c_g} Scale 2	PCQM $f8$
8	MS-GraphSIM SIM_{μ_g} Scale 2	PSNR MSE Y
9	PCQM $f7$	PSNR MSE U
10	PCQM $f8$	PSNR MSE V
11	PSNR MSE V	MS-GraphSIM SIM_{c_g} Scale 0
12	PointSSIM Geometry Features	MS-GraphSIM SIM_{m_g} Scale 1
13	PointSSIM Luminance Features	PointSSIM Geometry Features
14	MS-GraphSIM SIM_{μ_g} Scale 0	MS-GraphSIM SIM_{c_g} Scale 2
15	MS-GraphSIM SIM_{μ_g} Scale 1	PCQM $f1$
16	PSNR MSE U	MS-GraphSIM SIM_{μ_g} Scale 2
17	PSNR MSE Y	MS-GraphSIM SIM_{μ_g} Scale 0
18	MS-GraphSIM SIM_{c_g} Scale 1	MS-GraphSIM SIM_{m_g} Scale 2
19	MS-GraphSIM SIM_{m_g} Scale 1	MS-GraphSIM SIM_{μ_g} Scale 1
20	PCQM $f3$	MS-GraphSIM SIM_{c_g} Scale 1
21	PCQM $f6$	PointSSIM Luminance Features
22	MS-GraphSIM SIM_{c_g} Scale 0	PCQM $f6$
23	PCQM $f1$	PCQM $f3$



(a) *p13* (Basics) (b) *p13* V-PCC (c) *p13* G-PCC Predlift (d) *p13* G-PCC RAHT (e) *p13* GeoCNN

(f) *p24* (g) *p24* V-PCC (h) *p24* G-PCC Predlift (i) *p24* G-PCC RAHT (j) *p24* GeoCNN

(k) *p72* (l) *p72* V-PCC (m) *p72* G-PCC Predlift (n) *p72* G-PCC RAHT (o) *p72* GeoCNN

Figure 3.7: Examples of point clouds in the BASICS Database. The first column shows the reference point cloud, and the remaining ones depicts the lowest rate that results from each codec.

Table 3.5: Metric performance using ten-fold cross validation using the BASICS training dataset.

Metric Combination	**Regression Method**	*Features*	*Type*	***PCC***	σ_{PCC}	***SROCC***	σ_{SROCC}
Model 1 (8 features)	SVR	PCQM ($f2$, $f4$, $f5$, $f6$), MS-GraphSIM (SIM_{m_g} Scale 0, SIM_{μ_g} Scale 0, SIM_{c_g} Scale 0) PSNR MSE D2	FR $GEO+COL$	0.939	0.018	0.873	0.026
Model 2 (10 features)	SVR	PCQM ($f2$, $f4$, $f5$, $f6$, $f7$), MS-GraphSIM (SIM_{m_g} Scale 0, SIM_{c_g} Scale 0), PSNR MSE D2, PointSSIM Geometry and Luminance features	FR $GEO+COL$	0.939	0.022	0.886	0.043
Model 3 (14 features)	SVR	PCQM ($f2$, $f4$, $f5$, $f7$, $f8$), MS-GraphSIM (SIM_{m_g} Scale 0, SIM_{μ_g} Scale 0, SIM_{m_g} Scale 0, SIM_{μ_g} Scale 2, and SIM_{c_g} Scale 2), PSNR MSE D2, PSNR MSE V, PointSSIM Geometry and Luminance features,	FR $GEO+COL$	*0.949*	*0.013*	*0.892*	0.022
Model 4 (4 features)	SVR	PCQM ($f2$, $f4$, $f5$), MS-GraphSIM (SIM_{m_g} Scale 0)	FR $GEO+LUM$	0.944	0.022	0.878	0.03
Model 5 (6 features)	RR	PCQM ($f2$, $f4$, $f5$, $f7$), MS-GraphSIM (SIM_{m_g} Scale 0), PSNR MSE D2	FR $GEO+COL$	0.932	0.025	0.869	0.035
Model 6 (11 features)	RR	PCQM ($f2$, $f4$, $f5$, $f7$, $f8$), MS-GraphSIM (SIM_{m_g} Scale 0, SIM_{c_g} Scale 0, SIM_{m_g} Scale 2, SIM_{c_g} Scale 2), PSNR MSE D2, PointSSIM Geometry Features	FR $GEO+COL$	0.944	0.024	0.887	*0.016*
Model 7 (15 features)	RR	PCQM ($f1$,$f2$,$f4$,$f5$,$f7$,$f8$), MS-GraphSIM (SIM_{m_g} Scale 0, SIM_{c_g} Scale 0, SIM_{c_g} Scale 1, SIM_{c_g} Scale 2), PSNR MSE D2, PSNR MSE Y, PSNR MSE U, PSNR MSE V, PointSSIM Geometry Features	FR $GEO+COL$	**0.951**	**0.01**	**0.9**	**0.014**
Model 8 (4 features)	RR	PCQM ($f2$, $f4$, $f5$), MS-GraphSIM (SIM_{m_g} Scale 0)	FR $GEO+LUM$	0.927	0.027	0.87	0.04

as it is one the largest annotated quality assessment dataset containing artifacts generated solely by point cloud coding solutions.

3.2.3 Feature combination models

Once the most significant features were determined, every model was trained using combinations of features based on the ranking provided by the RFE, with the respective regression method.

The database was randomly partitioned at a ratio of 80%:20% for SVR training and testing, respectively. Data splitting was done at the level of the reference point clouds. Hence, it was assured that the reference or distorted versions of the same point cloud were either on the testing set or on the training set. After logistic fitting, the PCC and SROCC were computed for each of the ten random partitions.

During ten-fold cross validation, the mean squared error (MSE), mean absolute error (MAE) and coefficient of determination (r^2) were computed between the training scores and the test scores. Then, after the cross validation, we computed the mean for each indicator [128]. The results showed no case of overfitting, as the MSE, MAE and r^2 test values were never extremely higher than the training results. The MSE, MAE and r^2 scores ratio between training and testing set were always close to 1. Hence, it was concluded that the followed methodology did not lead to overfitting.

Furthermore, the standard deviation (σ) of both PCC and SROCC was also computed to

Table 3.6: Metric performance using the BASICS validation dataset. The metric combination models are defined in table 3.5

Metric	**Regression Method**	*Type*	***PCC***	***SROCC***	***RMSE***	***OR***	Average Time (s)
Model 1 (8 features)	SVR	FR, $GEO + COL$	0.936	**0.881**	0.098	0.717	66.324
Model 2 (10 features)	SVR	FR, $GEO + COL$	0.924	0.846	0.106	0.727	86.724
Model 3 (14 features)	SVR	FR, $GEO + COL$	0.933	0.845	0.100	0.683	109.312
Model 4 (4 features)	SVR	FR, $GEO + LUM$	0.937	0.840	0.097	0.707	43.736
Model 5 (6 features)	RiR	FR, $GEO + LUM$	**0.944**	0.854	**0.092**	0.670	66.32
Model 6 (11 features)	RiR	FR, $GEO + COL$	0.936	0.828	0.099	0.697	76.62
Model 7 (15 features)	RiR	FR, $GEO + COL$	0.938	0.840	0.097	0.690	99.208
Model 8 (4 features)	RiR	FR, $GEO + LUM$	0.930	0.832	0.103	0.687	43.732
PSNR MSE D1 [19]	-	FR, GEO	0.894	0.800	0.126	0.760	22.588
PSNR MSE D2 [19]	-	FR, GEO	0.923	0.836	0.108	0.693	22.588
PointSSIM Geometry Features [97]	-	FR, GEO	0.815	0.769	0.162	0.837	10.30
PointSSIM Luminance Features [97]	-	FR, LUM	0.718	0.677	0.194	0.840	10.10
PSNR MSE Y [95]	-	FR, LUM	0.580	0.550	0.229	0.903	22.588
PSNR MSE YUV [95]	-	FR, COL	0.638	0.567	0.215	0.907	22.588
Color Histogram [100]	-	FR, COL	0.497	0.428	0.244	0.883	0.015
PCQM [11]	-	FR, $GEO + LUM$	0.927	0.849	0.105	0.690	14.43
Point 2 Distribution [98]	-	FR, $GEO + COL$	0.748	0.612	0.186	0.847	24.524
GraphSIM [37]	-	FR, $GEO + COL$	0.924	0.817	0.108	**0.663**	46.88
MS GraphSIM [12]	-	FR, $GEO + COL$	0.909	0.808	0.117	0.710	29.30
PCM_{RR} [21]	-	RR, $GEO + COL$	0.567	0.493	0.232	0.860	124.28
RR_{CAP} [129]	-	RR, $GEO + COL$	0.749	0.538	0.186	0.840	3.104
FRSVR [130]	SVR	FR, $GEO + COL$	0.862	0.797	0.142	0.807	4.474

understand if there were large variations between each split. The results are shown in Table 3.5. The table shows the combination of metrics and features that lead to the best results. The second column of the table defines if the metric is full-reference (FR), reduced-reference (RR), or no-reference (NR). Furthermore, it is also described if the metric considers only the geometry (GEO), color (COL), luminance (LUM), both geometry and color ($GEO + COL$) information or both geometry and luminance ($GEO + LUM$) information. COL means that the selected features use both chromatic and luminance information. Between them, there were slight performance changes, and the models achieved a similar correlation for both PCC and SROCC. It is noted that the best performance is achieved by model 7 considering features from PCQM, MS-GraphSIM, PSNR MSE D2, PSNR MSE YUV, and PointSSIM ($PCC/SROCC = 0.951/0.9$). It can also be observed that the standard deviations of the PCC and SROCC are quite small, indicating that each fold had a similar performance. Nonetheless, it is noted that the best performance is achieved by model 7 considering features from PCQM, MS-GraphSIM, PSNR MSE D2, PSNR MSE YUV, and PointSSIM ($PCC/SROCC = 0.951/0.9$), using RiR as a regression method. It is closely followed by model 3. That model considers features defined in the same metrics, but trained with an SVR.

Once the most significant features were determined, every model was trained using combinations of features based on the ranking provided by the RFE, with the respective regression method.

Table 3.6 shows the quality features combination models performance on the BASICS validation dataset, after training the models with the complete training dataset. The validation dataset uses 15 reference point clouds that are used to generate 300 distorted point

clouds, using the same codecs as the training dataset. Moreover, the results of the state-of-the-art metrics in this database are also shown for comparison purposes. The final column shows the average time each metric took to compute, for the BASICS dataset. As expected, the computational complexity grows if more features are added.

It can be observed that the best PCC value is obtained using model 5, that contains features defined in PCQM and MS-GraphSIM, and the PSNR MSE D2, using the RR regression algorithm, and the best SROCC is achieved by model 1, containing features defined in PCQM and MS-GraphSIM, and the PSNR MSE D2, using an SVR. The obtained RMSE and OR are also quite low, when compared to the other state-of-the-art-metrics, although the best OR value is obtained by GraphSIM. It is also observed that the obtained results are different from the obtained in the training dataset.

3.2.4 Evaluating Feature Combinations

Objective quality metrics should be validated using subjective quality evaluation results as ground truth. As such, the default implementation of the state-of-the-art full-reference metrics PSNR MSE D1, PSNR MSE D2, PSNR MSE Y, PSNR MSE YUV, Point 2 Distribution, PointSSIM, PCQM, GraphSIM, Color Histogram, MS-GraphSIM, FRSVR and the reduced-reference metrics PCM_{RR} and RR_{CAP} were selected. A number of subjective quality evaluations were conducted in order to validate the models. It was ensured that none of the point clouds present in those evaluations were present in the training set. Subjective quality evaluation 1 reported on a subjective evaluation [16] comparing V-PCC and G-PCC using the octree mode, the deep-learning solution RS-DLPCC [52] and the Draco[8] codec. Subjective quality evaluation 2, which is focused on learning-based coding solutions [13] quality evaluation, was also chosen. In particular, the learning-based codecs ADLPCC [53], GeoCNNv2 [7], and PCGCv2 [6] performance was analyzed. The LUT_SR [34] solution was also considered. The codecs were compared to the octree mode of G-PCC. Finally, subjective quality evaluation 3, reports on the results of the JPEG Pleno Call for Proposals [10]. The call focused on deep-learning solutions. As the call is anonymous, the solutions are referred to as T1, T2, and T3, and they all can encode both geometry and color information of point cloud static content. The codecs were compared against G-PCC and V-PCC.

All chosen subjective quality evaluations consider deep-learning based solutions. This is important as these types of codecs are becoming increasingly popular, and extensive research is being conducted on them. For that reason, it is important to verify how the objective quality evaluation models perform when evaluating the performance of these codecs. Furthermore, two popular point cloud quality assessment (PCQA) datasets were considered, notably the Waterloo [89] and the Shanghai Jiao Tong University point cloud quality assessment (SJTU-PCQA) [90]. The Waterloo database contains 700 point clouds coded

[8] https://github.com/google/draco

with V-PCC, G-PCC (octree and trisoup modes), Downsampling and Gaussian noise, targeting both texture and geometry distortions. SJTU-PCQA contains 378 point clouds coded using octree-based compression, color noise, downscaling, downscaling and color noise, downscaling and geometry gaussian noise, geometry gaussian noise, color noise, and geometry gaussian noise. For the considered metrics and feature combination mod-

Table 3.7: Metrics performance for the datasets referred to as subjective evaluations 1, 2 and 3. The metric combination models are defined in table 3.5

		Subjective quality evaluation 1 [16]				Subjective quality evaluation 2 [13]				Subjective quality evaluation 3 [10]			
	Tested codecs	V-PCC, G-PCC, RS-DLPCC, Draco				G-PCC, ADLPCC,PCC GEO CNNv2, PCGCv2,LUT_SR				V-PCC, G-PCC, T1, T2, T3			
Metric	**Regression Method**	***PCC***	***SROCC***	***RMSE***	***OR***	***PCC***	***SROCC***	***RMSE***	***OR***	***PCC***	***SROCC***	***RMSE***	***OR***
PSNR MSE D1 [19]	-	0.890	0.884	0.148	0.618	0.806	0.782	0.184	0.753	0.741	0.725	0.226	0.781
PSNR MSE D2 [19]	-	0.851	0.847	0.169	0.608	0.821	0.796	0.177	0.813	0.783	0.773	0.210	0.769
PSNR MSE Y [95]	-	0.770	0.772	0.205	0.688	0.627	0.617	0.242	0.767	0.828	0.808	0.190	0.719
PSNR MSE YUV [95]	-	0.670	0.679	0.240	0.719	0.636	0.658	0.240	0.820	0.830	0.806	0.188	0.744
Color Histogram [95]	-	0.890	0.897	0.135	0.631	0.832	0.830	0.172	0.733	0.680	0.641	0.249	0.831
PCQM [11]	-	0.944	0.928	0.106	0.480	0.899	0.903	0.137	0.573	0.873	0.826	0.167	0.694
Point 2 Distribution [98]	-	0.778	0.794	0.204	0.747	0.851	0.828	0.164	0.640	0.866	0.833	0.169	0.794
PCM_{RR} [21]	-	0.890	0.871	0.147	0.529	0.834	0.834	0.172	0.727	0.837	0.831	0.185	0.763
RR_{CAP} [129]	-	0.718	0.685	0.226	0.833	0.735	0.734	0.212	0.867	0.813	0.822	0.197	0.675
PointSSIM [97]	-	0.869	0.867	0.160	0.588	0.859	0.857	0.159	0.720	0.706	0.684	0.239	0.831
GraphSIM [37]	-	0.907	0.893	0.137	0.500	0.800	0.799	0.186	0.780	0.919	0.900	0.135	0.569
MS-GraphSIM [12]	-	0.902	0.880	0.179	0.490	0.890	0.884	0.142	0.620	**0.925**	**0.901**	**0.130**	**0.500**
FRSVR [130]	SVR	0.811	0.763	0.189	0.686	0.780	0.655	0.194	0.753	0.679	0.661	0.247	0.838
Model 1 (8 features)	SVR	0.943	0.917	0.107	0.589	0.903	0.857	0.133	0.693	0.906	0.860	0.143	0.675
Model 2 (10 features)	SVR	0.945	0.913	0.105	0.520	0.889	0.834	0.143	0.720	0.907	0.855	0.143	0.688
Model 3 (14 features)	SVR	0.938	0.871	0.112	0.529	0.884	0.814	0.145	0.753	0.905	0.853	0.144	0.681
Model 4 (4 features)	SVR	0.961	0.937	0.089	0.490	0.924	0.888	0.118	0.707	0.906	0.850	0.143	0.675
Model 5 (6 features) (FSM)	RiR	**0.958**	**0.939**	**0.093**	0.490	0.916	0.908	0.123	0.653	0.909	0.878	0.142	0.625
Model 6 (11 features)	RiR	0.951	0.937	0.100	0.510	0.919	0.913	0.120	0.640	0.897	0.872	0.150	0.600
Model 7 (15 features)	RiR	0.947	0.935	0.104	0.480	0.913	0.907	0.124	0.633	0.904	0.880	0.145	0.613
Model 8 (4 features)	RiR	0.956	0.938	0.095	**0.422**	**0.930**	**0.926**	**0.113**	**0.553**	0.894	0.866	0.153	0.606

Table 3.8: Metrics performance for Waterloo and SJTU-PCQA. The metric combination models are defined in table 3.5

		Waterloo [89]				SJTU-PCQA [90]			
Metric	**Regression Method**	***PCC***	***SROCC***	***RMSE***	***OR***	***PCC***	***SROCC***	***RMSE***	***OR***
PSNR MSE D1 [19]	-	0.578	0.566	0.203	0.935	0.873	0.798	0.135	*0.802*
PSNR MSE D2 [19]	-	0.481	0.461	0.219	0.932	0.762	0.678	0.180	0.830
PSNR MSE Y [95]	-	0.608	0.587	0.197	0.939	0.701	0.704	0.197	0.892
PSNR MSE YUV [95]	-	0.551	0.536	0.207	0.935	0.655	0.659	0.211	0.923
Color Histogram [95]	-	0.195	0.199	0.243	0.951	0.068	0.111	0.280	0.934
PCQM [11]	-	0.750	0.743	0.165	0.884	0.858	0.844	0.142	0.812
Point 2 Distribution [98]	-	0.462	0.432	0.222	0.932	0.632	0.620	0.217	0.881
PCM_{RR} [21]	-	0.368	0.345	0.232	0.931	-	-	-	-
RR_{CAP} [129]	-	0.708	0.715	0.176	0.936	0.765	0.752	0.180	0.899
PointSSIM [97]	-	0.468	0.455	0.220	0.928	0.723	0.705	0.191	0.910
GraphSIM [37]	-	0.690	0.681	0.180	0.918	0.868	0.854	0.138	0.820
MS-GraphSIM [12]	-	0.716	0.708	0.174	0.914	**0.893**	**0.874**	**0.125**	0.831
FRSVR [130]	SVR	0.391	0.181	0.228	0.949	0.606	0.614	0.220	0.902
Model 1 (8 features)	SVR	0.676	0.686	0.183	0.892	0.872	0.856	0.136	0.836
Model 2 (10 features)	SVR	0.676	0.680	0.183	0.908	0.879	0.859	0.132	0.839
Model 3 (14 features)	SVR	0.679	0.688	0.183	0.916	0.889	0.865	0.127	**0.786**
Model 4 (4 features)	SVR	0.758	0.760	0.162	0.896	0.882	0.840	0.130	0.815
Model 5 (6 features) (FSM)	RiR	0.702	0.715	0.177	0.896	0.889	0.870	0.127	0.825
Model 6 (11 features)	RiR	0.686	0.681	0.247	0.947	0.856	0.842	0.143	0.841
Model 7 (15 features)	RiR	0.686	0.690	0.181	0.912	0.874	0.857	0.134	0.815
Model 8 (4 features)	RiR	**0.788**	**0.790**	**0.153**	**0.884**	0.881	0.868	0.131	0.823

els, the statistical measures proposed in [131] were computed, specifically the PCC, the SROCC, the Root Mean Squared Error (RMSE) and the Outlier Ratio (OR). The MOS predicted for each of the objective metrics was calculated by applying a logistic fit function to the objective scores, as is commonly done when benchmarking objective metrics [114, 115]. All the combinations in table 3.5 were evaluated, as they show similar performance. The results are shown in tables 3.7 and 3.8.

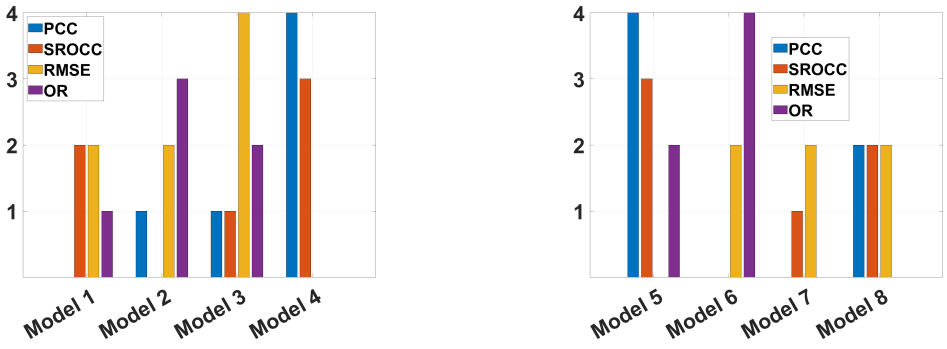
Analyzing the results for the subjective quality evaluation 1 [16], it can be observed that the defined models that combine several features achieve better performance than the individual metrics. It should be noted that the models have a very similar performance for both PCC and SROCC as well as RMSE and OR, independently of the regression method. The best performing model for this evaluation is model 5 (6 features with RiR).

Regarding subjective quality evaluation 2 [13], it can be observed that models 8 (4 features with RiR) achieves the best correlation values.

Furthermore, both RMSE and OR are also lower than the ones obtained for the state-of-the-art metrics for each combination model.

The subjective quality evaluation 3 [10] provided very competitive results, although the performance provided by GraphSIM and MS-GraphSIM is slightly better. In this evaluation, the performance achieved by the different models is once again very similar, although the model 5 (6 features trained with RiR) achieves the best performance.

The results for the Waterloo and SJTU-PCQA dataset, shown in table 3.8, reveal a lower performance. Moreover, other metrics also have a reduced performance for both databases. These databases contain a wide range of distortions including gaussian noise and down-scalling, not present in the BASICS database (used for training) which causes this performance reduction. However, even in this situation, the combination models obtained for the different regression methods generally achieve the best performance. For the Waterloo database, model 8 (4 features with RiR) shows the best performance. For the SJTU-



(a) SVR

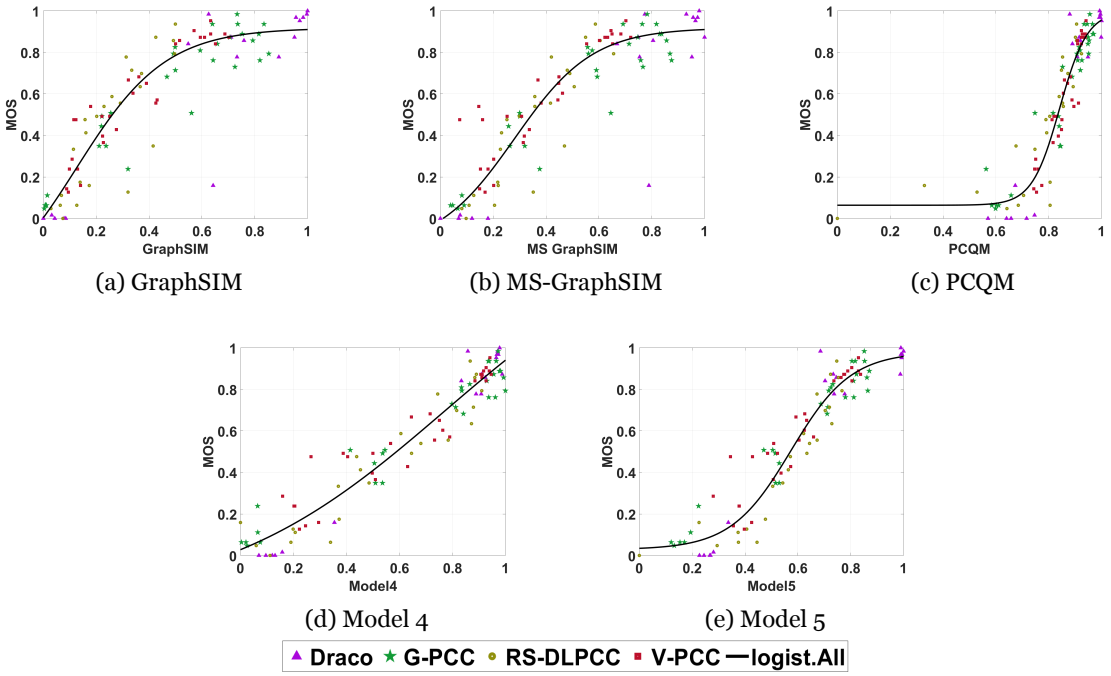
(b) RiR

Figure 3.8: Histograms representing the number of times that each model performs the best for PCC, SROCC, RMSE and OR. (Figure taken from [15].)

PCQA dataset, the best performing metric is the MS-GraphSIM, closely followed by model 5 (6 Features with RiR).

From the results shown in tables 3.5, 3.7 and 3.8, it is very difficult to select the best performing model. Most of the combination models achieved very high performance for PCC, SROCC, RMSE and OR. To assess the most consistent ones, the times that a model achieved the best PCC, SROCC, RMSE or OR values were identified, allowing to access the best model for each regression method. Then the same comparison was conducted between the identified best models. Fig. 3.8 shows the results for each feature combination model, and shows that models 4 and 5 are the best models for the SVR and RiR regression methods, respectively.

Fig. 3.9 shows the plots of normalized MOS vs. metrics for each of the best models and the best metrics for the subjective quality evaluation 1. It can be observed that the results for each model are close to the logistic curve, but the curve of model 5 reveals to be the best fit.



(a) GraphSIM (b) MS-GraphSIM (c) PCQM

(d) Model 4 (e) Model 5

Figure 3.9: Metric vs. MOS plots, with logistic regression curves of FSM and the three best performing metrics for subjective quality evaluation 1. [16].

3.3 Normal Computation methods

Some metrics require the point cloud normal vectors. A study conducted by the JPEG Pleno Learning Based Point Cloud Coding committee identified two relevant methods for

computing those normals, mainly using quadric fitting [132] of CloudCompare [9], and the nearest neighbors of MeshLab. A comparison between the two point cloud normals computation methods was reported in Prazeres *et al.* [5]

The quadric fitting method starts by defining a local neighborhood for points, for each point of the point cloud. That neighborhood is defined by a sphere with a fixed radius. Within the neighborhood, a quadric surface if fitted to the points, as defined by equation 3.3.

$$ax^2 + by^2 + cz^2 + 2dxy + 2exy + 2fyz + 2gx + 2hy + 2iz + j = 0 \quad (3.3)$$

where the coefficients a to j are determined by minimizing the fitting error. Once the quadric surface is fitted to the local neighborhood, the normal vector at the point of interest is derived from the gradient of the fitted surface. The gradient gives the direction perpendicular to the fitted surface, which is the normal.

The k-NN algorithm in Meshlab identifies the closest k points based on the Euclidean distance. After identification of the k closest neighbors, the covariance matrix of these neighborhood points is computed. It is then subjected to eigenvalue decomposition. The normal vector is defined as the eigenvector with the smallest eigenvalue.

In a study developed during this doctoral program [5], the influence of these methods was assessed. Cloud Compare quadric fitting was employed with a radius of 5, 10 and 20, while Meshlab k-NN was computed with a k of 6, 10 and 16. The values were selected in concordance with a previous study [133]. Tables 3.9, 3.10 and 3.10 show the results obtained for three different metrics for two distinct datasets, namely the EI2022 [16] and the BASICS dataset [18]. It can be observed that the Quadric Fitting algorithm leads to a superior performance than k-NN from Meshlab.

3.4 Statistical Analysis

In order to assess the differences between subjective and objective evaluations, statistical tests must be conducted. This reveals if there are any statistically relevant differences between subjective evaluation protocols, or in the behavior of objective quality metrics. For subjective quality evaluations, either the Kruskal-Wallis test [134] or an Analysis of variance (ANOVA) [135] can be considered. For objective quality metrics, usually the Krasula Method is considered [136].

[9] *https://cloudcompare.org/doc/wiki/*

Table 3.9: Influence of the plane estimation method on PSNR MSE D2 [19] metric

	EI2022 [16]				BASICS [18]			
Method	***PCC***	***SROCC***	***RMSE***	***OR***	***PCC***	***SROCC***	***RMSE***	***OR***
Quadric R =5	0.828	0.827	0.181	0.618	0.919	0.836	0.110	0.710
Quadric R =10	0.834	0.831	0.179	0.608	0.920	**0.838**	0.109	0.723
Quadric R =20	**0.851**	**0.847**	**0.169**	0.608	**0.923**	0.836	**0.108**	**0.693**
KNN K =6	0.806	0.792	0.192	0.618	0.919	**0.838**	0.110	0.710
KNN K =10	0.810	0.805	0.190	**0.598**	0.920	**0.838**	0.110	0.717
KNN K =18	0.816	0.811	0.187	0.598	0.921	**0.838**	0.109	0.717

Table 3.10: Influence of the plane estimation method on PL2Plane [20] metric.

	EI2022 [16]				BASICS [18]			
Method	***PCC***	***SROCC***	***RMSE***	***OR***	***PCC***	***SROCC***	***RMSE***	***OR***
Quadric R =5	0.806	0.792	0.191	0.735	**0.807**	**0.738**	**0.166**	**0.817**
Quadric R =10	0.827	0.771	0.182	0.725	0.750	0.649	0.185	0.850
Quadric R =20	**0.846**	**0.795**	**0.172**	**0.667**	0.650	0.628	0.212	0.877
KNN K =6	0.318	0.297	0.309	0.892	0.453	0.457	0.250	0.880
KNN K =10	0.387	0.357	0.301	0.863	0.451	0.502	0.250	0.863
KNN K =18	0.473	0.373	0.286	0.833	0.434	0.493	0.253	0.893

Table 3.11: Influence of the plane estimation method on PCM_{RR} [21] metric.

	EI2022 [16]				BASICS [18]			
Method	***PCC***	***SROCC***	***RMSE***	***OR***	***PCC***	***SROCC***	***RMSE***	***OR***
Quadric R =5	0.884	0.872	0.151	0.588	**0.639**	**0.549**	**0.215**	**0.840**
Quadric R =10	**0.890**	**0.871**	**0.147**	**0.529**	0.612	0.516	0.221	0.867
Quadric R =20	0.842	0.843	0.174	0.559	0.594	0.493	0.224	0.883
KNN K =6	0.670	0.653	0.242	0.784	0.608	0.576	0.221	0.857
KNN K =10	0.673	0.642	0.240	0.765	0.617	0.581	0.219	0.847
KNN K =18	0.781	0.767	0.201	0.667	0.616	0.565	0.219	0.857

3.4.1 Kruskal-Wallis

This is a non-parametric test, evaluating if a certain group of samples originates from the same distribution. The samples under comparison can have either equal or different sizes. This method can be used to evaluate more than one group, making it an extension of the Mann–Whitney U test [137] and it is the non-parametric alternative to the ANOVA Test [135].

The test starts by combining and group data and ranking the values. Then the sum of ranks of each group is computed. Then the Test Statistic (H) is computed using:

$$H = (\frac{12}{N(N+1)}) \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \quad (3.4)$$

where N is the total number of observations, k is the number of groups, R_i is the sum of the ranks for the i-th group and n_i is the number of observations in the i-th group. The H value is then compared to the χ^2 (chi-square) distribution with k-1 degrees of freedom to

determine the p-value. Then, if the p-value is less than 0.05, it is decided that the samples come from different distributions.

3.4.2 Multi-way ANOVA Test with repeated measurements

This test is an extension of the one-way ANOVA, that allows to evaluate the influence of two or more independent variables, on a dependent variable. It is used to examine the main effects of each factor or the interaction effects between each factor. For this evaluation, it is assumed that the variables are normally distributed.

Firstly, the variability of the data is computed, followed by a computation of the mean squares. Then, the F-statistic is computed using the following equation:

$$F = \frac{MS_{effect}}{MS_{error}} \quad (3.5)$$

The F statistic is then compared to critical F-values, which are the threshold for determining statistical difference (usually below 0.05).

3.4.3 Krasula Method

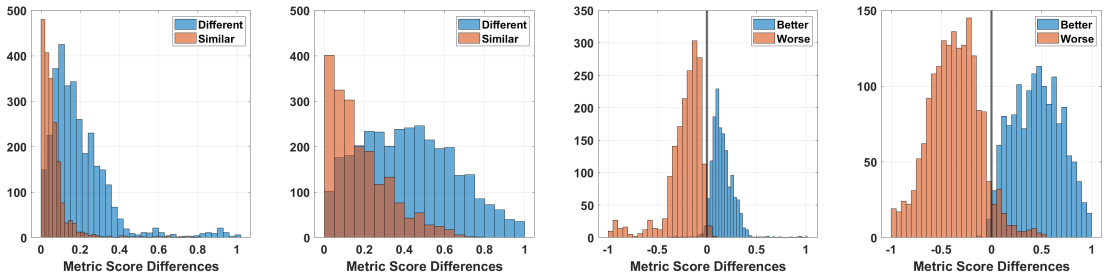
This method evaluates the performance of objective metrics in two different analysis, notably ”Different vs Similar” and ”Better vs Worse”.

In the first analysis, the distorted point clouds are paired and split into two different categories (Different and Similar). The ”Different” category represents pairs with statistically significant differences, and the ”Similar” the ones without. For each pair, a one way ANOVA followed by a Turkeys honest significance difference test [138] is conducted, measuring the statistical significance of the differences. The Krasula method assumes that the absolute difference of predictions made by the metrics for different pairs should be larger than the similar pairs.

To quantify the performance, the Receiving Operating Characteristic (ROC) analysis is used, and the performance is expressed as the Area Under ROC Curve (AUC).

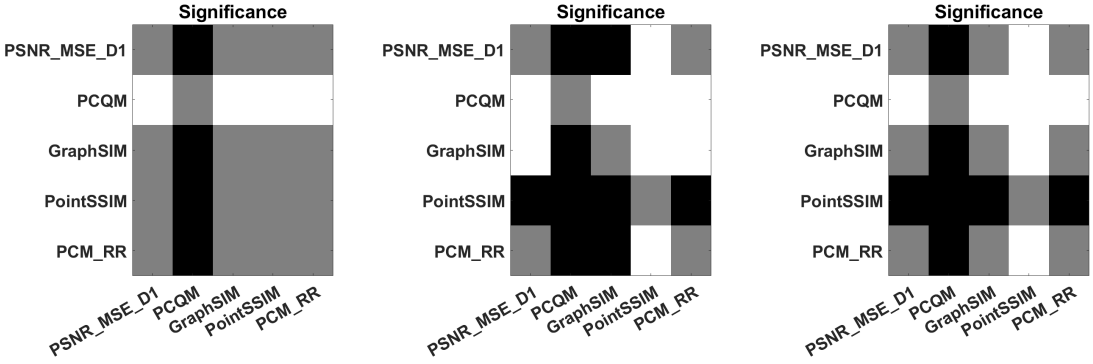
The second analysis (Better vs Worse) considers only the pairs that have statistically significant differences, through the evaluation the performance of the metrics based on the identification of the better pairs with statistically significant differences. The performance is evaluated as correct classification percentage (CCp), and AUC. Fig. 3.10 shows an example for two metrics, namely PCQM [11] and GraphSIM [37] for the EI2022 dataset [16].

Finally, the Krasula method also performs statistical significance tests. Notably, it employs the Hanley-Macneil Method [139], to compare AUC values from ROC analysis. Furthermore, the Fisher’s [140] exact test is conducted to compare the percentage of correct



(a) 1 - PCQM (b) GraphSIM (c) 1 - PCQM (d) GraphSIM

Figure 3.10: Depiction of Different vs Similar and Better vs Worse (figure taken from [5]).



(a) Different vs Similar (AUC) (b) Better Worse (AUC) (c) Better Worse (CCp)

Figure 3.11: Statistical Analysis results for the EI2022 Dataset [16] (figure taken from [5]).

recognition of the stimulus of higher quality. Fig. 3.11 shows an example of a significance test for the five best performing metrics.

3.5 Analyzing Learning-based codecs performance

When dealing with learning-based methods, it is well known that the performance of the final learned model may vary, even with similar training conditions, due to the stochastic nature of the learning process [32].

Table 3.12: BD-Metrics and BD-Rate using G-PCC as a reference.

	PSNR MSE D1		PSNR MSE D2		PCQM	
Codecs	**BD-PSNR**	**BD-Rate**	**BD-PSNR**	**BD-Rate**	**BD-Metric**	**BD-Rate**
VmUC Train 1	2.908	-26.302	2.204	-18.773	-1.746E-03	67.724
VmUC Train 2	3.369	-56.292	2.876	-51.759	-1.705E-03	48.335
VmUC Train 3	3.149	-56.964	2.660	-46.735	-2.240E-03	60.528
VmUC Default	4.359	-72.885	4.235	-66.397	-1.603E-03	48.193
V-PCC	3.128	-60.617	1.695	-26.307	1.406E-03	-27.885

To evaluate how the compression quality evolves during training, as well as how the final epochs differ, for each step of the learning progression, a point cloud is encoded and the

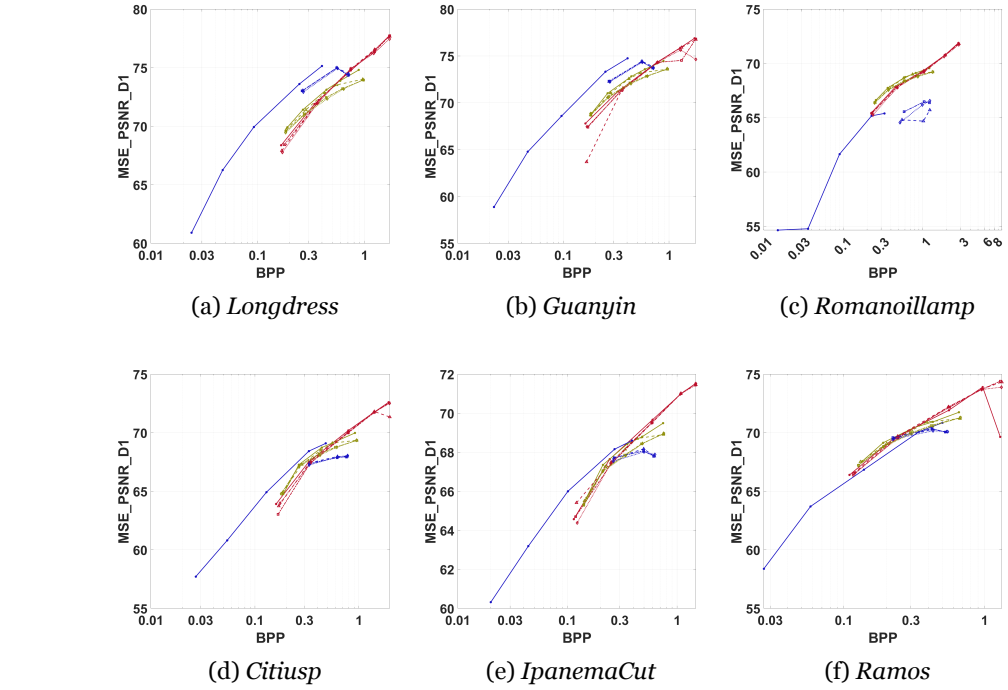
point to point metric (PSNR MSE D1) [19] is computed. The PSNR MSE D1 metric is usually selected for quality estimation of codecs that only encode geometry. Furthermore, it is usually the best geometry only metric in benchmarking studies [81, 16, 13]. The obtained results are then analyzed to verify if there is a convergence to a stable operating point. If the codec is able to encode color information, such as the work proposed by Guarda *et al.* [8], the PCQM metric and the GraphSIM metric should be used [36], as they usually provide a good representation of the subjective quality scores [5]. Table 3.12 shows the BD Rates obtained in a study conducted during this doctoral program [36]. Three different training sessions of the JPEG Pleno VM were conducted, and the results were compared to G-PCC. Furthermore, the default implementation of the JPEG Pleno Learning based Point Cloud VM, and V-PCC are also included. It can be observed that different training sessions can lead to very different results.

Another study was conducted on the stability of learning-based codecs, that encode only geometry, namely ADLPCC, PCC GEO CNNv2 and PCGCv2 [124]. The procedure followed the same method as the previous work, with the addition of a comparison with the default codec, shown in figs 3.12 and 3.13.

From the figures, it can be concluded that ADLPCC depends less on the training session, but has a lower performance on high bitrates, in some cases. Higher bit rates are the ones that provide an acceptable quality, and they might be the ones most commonly used in practical applications, which makes this inconsistency a problem.

PCC GEO CNNv2 has one training situation that consistently leads to much better performance in the middle bit rates than the other training situations. This level of variation establishes some unreliability in the codecs performance.

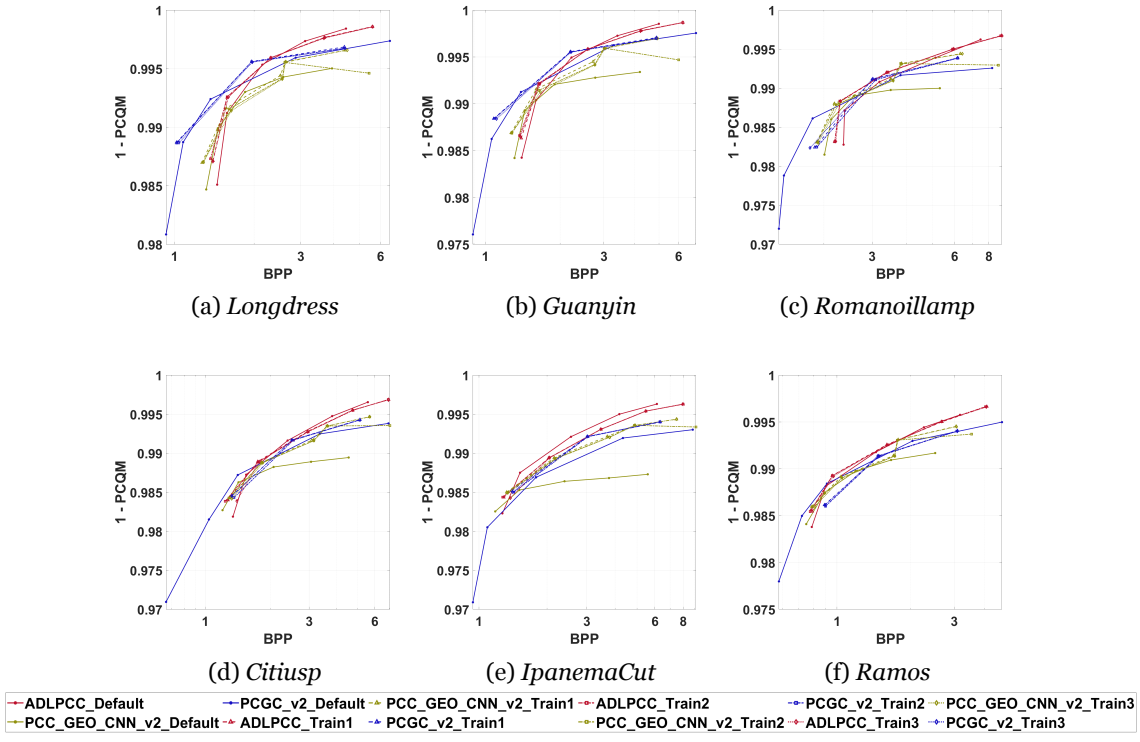
PCGCv2 is revealed to be the most stable codec, with the only exception being the *Romanoillamp* point cloud. It should be noted that the authors do not specify the lambda trade-off that they use in the current implementation. As such, some results will vary depending on the default codec and the training sessions conducted.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp*

(d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Figure 3.12: MSE PSNR D1 plots for each of the defined operating points for each codec.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp*

(d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Figure 3.13: 1 - PCQM plots for each of the defined operating points for each codec.

Chapter 4

Conclusions and Future Work

The main goal of this thesis was to study, analyze, and develop quality models to evaluate the performance of point cloud coding solutions. Furthermore, the typical subjective evaluation protocols were tested on immersive environments to verify if they were suitable for subjective quality assessment in virtual reality or 3D stereoscopic displays. This chapter presents the conclusions that result from the research work conducted during this doctoral program. The chapter also discusses additional research topics.

4.1 Final Discussion

4.1.1 Subjective Quality Assessment

This doctoral program allowed to analyze the performance and stability of the subjective quality models developed by the JPEG committee for the evaluation of point cloud coding solutions, considering the development of a standardized JPEG Pleno Point Cloud codec. The subjective models were revealed to be stable, independent of the viewing display technology.

Furthermore, this thesis allowed to shed insight on conducting subjective quality assessment in immersive environments. While there are some works in the literature that conduct such quality studies in such environments, there were no works that conducted a direct comparison between studies using the same codecs, the same content, but different environments. The comparison between the developed works allowed the conclusion that the designed visualization model using HMD and a stereoscopic display resulted in a subjective evaluation that provided a very similar quality representation to the one that resulted using the common model using a 2D display [14, 31].

The subjective quality evaluation protocol using HMD can also be used with other displaying technology, avoiding the use of a large number of resources, notably the generation of 4k videos with lossless compression for each distorted content.

The quality evaluation of codecs that only encode geometry information is problematic. Subjective quality evaluation of the geometry only codecs, has revealed to be very unstable. Hence, there is a need to add the texture information to the encoded geometry of the point clouds to use the usual subjective quality evaluation model. Studies conducted during this doctoral program allowed to conclude that the most reliable method for subjec-

tive quality evaluation is to map the texture of the reference point cloud and then encode the texture with G-PCC (lossless for geometry and lossy for attributes mode) [124]. This methodology also provided the best representation of the objective quality metrics. In fact, with this model, typical objective metrics provide the expected level of quality representation, in opposition to other methodologies, like subjective evaluation of the geometry only (without texture information), or direct mapping of the texture.

4.1.2 Objective Quality Assessment

Regarding objective point cloud quality metrics, the main conclusion that can be drawn from the conducted work is that they struggle to provide an acceptable performance when predicting the quality of point clouds coded with learning-based technologies [5, 10, 13]. The conducted studies using learning-based solutions revealed that most metrics always achieve Pearson and Spearman correlations below 0.9. The only metrics that achieved acceptable results were the PCQM, GraphSIM, and the later multiscale counterpart, MS-GraphSIM. Furthermore, the conducted work allows the research community to understand what the best metrics are to consider when benchmarking point cloud coding solutions.

During this doctoral program the quality representation provided by different objective quality metric features was analyzed. The study allowed to build a model using the best combination of objective metrics and features that presented a very good representation of the quality in multiple databases with subjective scores of coded content.

Another conclusion regarding point cloud metrics is that the no-reference metrics are not yet suitable for a reliable quality evaluation [5]. Most metrics are trained in databases that contain a great deal of artifacts not usually present in the typical point cloud codecs. For this reason, the metrics show good results in certain point cloud databases but fail to provide acceptable outcomes in a general subjective evaluation. Furthermore, in some cases, the datasets in which the metrics are trained, only a subset is annotated using subjective evaluation. The MOS for the remaining point clouds is computed using that MOS as a base and leads to unreliable results.

One limitation of the no-reference metrics is that, since most of them use deep-learning technology, data-augmentation models cannot be employed. Some of the changes created by these augmentation models are likely to cause changes in the subjective quality, which will lead to unreliable training data.

4.1.3 Learning based point cloud coding solutions

It was verified that learning-based solutions are quickly surpassing the traditional codecs. Even the early solutions show some gains over both G-PCC and V-PCC when encoding the

geometry information of point clouds [13], and that superiority has been increasing over the years [10]. Even so, the learning-based solutions struggle to achieve superior results compared to V-PCC and G-PCC when encoding the texture attributes present in point clouds [10, 5, 16].

It is also revealed that learning-based point cloud codecs are very dependent on the training [32, 36, 124]. Studies show that different training sessions will lead to different working points, and sometimes the default implementation of the codec may not be the most efficient one.

4.2 Future Work

Quality assessment of point cloud coding solutions is a field that will continue to pose challenges and is expected to undergo significant developments. One promising research path is exploring different methods to present the content of subjective evaluation to the test subjects. One such method that was not explored during this doctoral program was the pairwise comparison methodology, where the same quality level between codecs is compared. Instead of a score, the subjects are asked to answer which one they prefer. This model has been recently proven to be very effective for high-quality images and for light field subjective quality assessment.

These models have been proving very appropriate for high-quality assessment in domains where the direct comparison used in this work is difficult to visualize.

Quality assessment of visually or near lossless point cloud coding solutions is again another field with limited exploration.

Finally, this work did not consider dynamic point clouds, mostly because of the involvement with JPEG. However, the subjective evaluation of dynamic point clouds requires new studies. Currently this evaluation is made using a visualization path through the point cloud sequence. However, these models should be validated using different visualization paths. That is not the case with the current state-of-the-art. Hence, new studies on the influence of the visualization path of the point cloud sequence and on the stability of the point cloud quality evaluation independent of the visualization path are required.

Chapter 5

Publications

5.1 Quality Evaluation of Point Cloud Compression Techniques

Quality evaluation of point cloud compression techniques
J. Prazeres, Manuela Pereira, Antonio M.G. Pinheiro
Signal Processing: Image Communication, Volume 128, 2024, 117156, ISSN 0923-5965
DOI:https://doi.org/10.1016/j.image.2024.117156

Signal Processing: Image Communication 128 (2024) 117156

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Quality evaluation of point cloud compression techniques

Joao Prazeres [a,b,*], Manuela Pereira [a,c], Antonio M.G. Pinheiro [a,b]

[a] Universidade da Beira Interior, Rua Marquês D'Ávila e Bolama, Covilha, 6201-001, Portugal
[b] Instituto de Telecomunicacoes, Rua Marquês D'Ávila e Bolama, Covilha, 6201-001, Portugal
[c] NOVA LINCS, Rua Marquês D'Ávila e Bolama, Covilha, 6201-001, Portugal

ARTICLE INFO

Keywords:
Point clouds
Quality evaluation
Coding
Machine learning

ABSTRACT

A study on the quality evaluation of point clouds in the presence of coding distortions is presented. For that, four different point cloud coding solutions, notably the standardized MPEG codecs G-PCC and V-PCC, a deep learning-based coding solution RS-DLPCC, and Draco, are compared using a subjective evaluation methodology. Furthermore, several full-reference, reduced-reference and no-reference point cloud quality metrics are evaluated. Two different point cloud normal computation methods were tested for the metrics that rely on them, notably the Cloud Compare quadric fitting method with radius of five, ten, and twenty and Meshlab KNN with K six, ten, and eighteen. To generalize the results, the objective quality metrics were also benchmarked on a public database, with mean opinion scores available. To evaluate the statistical differences between the metrics, the Krasula method was employed. The Point Cloud Quality Metric reveals the best performance and a very good representation of the subjective results, as well as being the metric with the most statistically significant results. It was also revealed that the Cloud Compare quadric fitting method with radius 10 and 20 produced the most reliable normals for the metrics dependent on them. Finally, the study revealed that the most commonly used metrics fail to accurately predict the compression quality when artifacts generated by deep learning methods are present.

1. Introduction

The current technological evolution is experiencing an increasing need for emergent 3D data formats. 3D information can be represented in multiple formats like point clouds, meshes, holograms, volumetric imaging, imaging slicing or indirectly using multi-view systems. Among these formats point clouds have emerged as one of the most popular methods.

Point clouds consist of a set of Cartesian coordinates (x, y, z) representing each point, with a list of attributes associated with each coordinate, such as an RGB component, reflective information, physical sensor information, or normal vectors. As with any 3D representation, point clouds are usually represented by huge amounts of data, allowing an extremely accurate representation of an object or scene, making them a very powerful representation model. Hence, most applications making use of this type of data can benefit from efficient coding solutions that provide the means for efficient processing, storage, and transmission. The main contributions of this paper are as follows:

- Report on a subjective quality evaluation using four different relevant codecs.
- Benchmarking of full-reference, reduced-reference and no-reference metrics.
- Analyze the influence of the normals computation methodology and parameterization.

This paper aims to report the knowledge gained on colored static point cloud quality evaluation under coding distortions, as well as the performance of objective point cloud quality metrics. Furthermore, up to the best of the authors knowledge, it is the first work that directly compares full-reference, reduced-reference and no-reference metrics, using two different datasets. The subjective quality study reported by Prazeres et al. [1] (referred to as EI2022 in the remainder of the paper) is expanded by increasing the number of evaluated metrics, and conducting more in-depth benchmarking. Furthermore, results are generalized using the BASICS database [2]. Additionally, more details on the subjective quality evaluation protocol, as well as visual examples are provided.

Six point clouds were carefully selected and encoded with four relevant codecs. The subjective quality evaluation methodology is described, and the performance of several commonly used objective evaluation metrics is analyzed under different types of compression artifacts. Moreover, the considered metrics are benchmarked on the BASICS validation dataset [2]. This is a large database that also contains

* Corresponding author at: Universidade da Beira Interior, Rua Marquês D'Ávila e Bolama, Covilha, 6201-001, Portugal.
E-mail address: joao.prazeres@ubi.pt (J. Prazeres).

https://doi.org/10.1016/j.image.2024.117156
Received 22 June 2023; Received in revised form 13 May 2024; Accepted 24 May 2024
Available online 7 June 2024

a deep-learning codec used for metrics validation. With this dataset, the results can be generalized, and better conclusions can be drawn. In particular, the study considers artifacts created by deep learning technology, which tend to be different from those created with octree or projection-based codecs.

After this introduction, the paper reports a short state-of-the art description of point cloud coding, as well as subjective and objective quality evaluation (Section 2). Afterward, the description of the quality study (Section 3), which includes the dataset selection, the subjective quality evaluation, the objective quality study, and metrics benchmarking, is presented. Finally, the results discussion (Section 4) provides some insights on the results of the conducted study and the final comments section (Section 5) summarizes the findings of the study, as well as providing recommendations for further studies.

2. Related work

2.1. Point cloud coding

The most traditional coding methods for point clouds are based on the octree pruning method [3]. Recently, MPEG defined the Geometry-based Point Cloud Compression (G-PCC) [4], for static point clouds based on the octree point cloud representation. G-PCC also allows the use of the trisoup method based on surface reconstruction for geometry compression. Regarding the point cloud attributes, they are compressed using either Region Adaptive Hierarchical Transform (RAHT) [5] or the predicting/lifting (predlift) transform [4].

Point cloud coding can also be achieved by encoding the point cloud projections, which can be coded by any image coding codec. MPEG also explored that approach, resulting in the Video-based Point Cloud Compression (V-PCC) [4] defined for dynamic point clouds. V-PCC relies on HEVC, or more recently, VVC, in order to encode 2D projections of a given point cloud. Despite being developed for dynamic point clouds, its intra-coding has been revealed to be the most efficient for static point cloud coding [1,6].

Following the good performance in image coding, several machine learning-based coding solutions for point clouds have been proposed recently [7–14]. The stability of machine-learning codecs was studied [15] using the PSNR D1 [16] metric to compute a set of point clouds across three training sessions conducted under the same circumstances. Machine learning solutions usually cause distortions represented by empty spaces in the point cloud geometry [17], which are quite different from those caused by common codecs. Hence, there is a need to analyze the reliability of the quality methodologies for learning-based codecs. For this study, four relevant point cloud coding solutions were selected.

The Video Point Cloud Compression (V-PCC) [4], uses HEVC to encode the generated projections and uses intracoding. The Geometry Point Cloud Compression (G-PCC) [4], with the octree method for geometry encoding and the lifting transform for texture encoding. Considering the work presented by Perry et al. [6], it was decided to use the predlift transform. The Resolution Scalable Deep Learning Point Cloud Compression (RS-DLPCC) [8] makes use of deep learning technology to compress the point cloud geometry. A latent representation of a point cloud is computed by an autoencoder framework. The interlaced block creation makes the scalability feature possible. The point cloud is divided into superblocks that are further divided by interlaced downsampling. This creates eight interlaced blocks for each defined superblock. The resulting blocks are then coded separately, enabling random access.

Draco[1] was developed by Google. The codec uses KD-Tree [18] in order to organize 3D data efficiently. The codec continuously splits the point cloud from the center, modifying the axes in each direction.

[1] https://github.com/google/draco

The four codecs represent the diversity of coding artifacts usually created by the lossy coding of point clouds. G-PCC is based on octree decomposition, V-PCC compresses the point cloud projections, and RS-DLPCC creates typical distortions that appear in deep learning-based codecs. Even though Draco was mainly developed for meshes, its performance on static [19] and dynamic point clouds has been studied [20,21]. For this reason, the codec was included in this evaluation.

2.2. Subjective quality evaluation

The research community has been extremely active in studying point cloud quality evaluation methodologies and conducting research on different coding methodologies and setups. Studies were conducted to establish quality methodologies for geometry-only methods [22,23], graph-based [24] and projection-based [25]. Honglei et al. conducted subjective evaluations on the MPEG test model V-PCC and also early proposals that led to the final version of G-PCC. A study on the MPEG codecs V-PCC and G-PCC before standardization was also conducted [26]. In 2020, a subjective and objective quality evaluation comparing the MPEG codecs was conducted [6], using a 2D setup. It was reported that V-PCC was the best performing codec. Recently, a subjective quality evaluation using 3D stereoscopic visualization [27] was compared with 2D visualization, revealing high correlations between both evaluations and no statistical differences. Moreover, a subjective quality evaluation targeting machine-learning-based coding solutions was reported [17]. In early 2022, a quality assessment study was performed to support the JPEG Pleno point cloud coding CfP [28]. This study aimed to evaluate the current state-of-the-art point cloud solutions, analyze the stability of the subjective quality assessment methodologies, and evaluate the performance of objective metrics. Subjective quality evaluation studies also lead to the definition of point cloud quality assessment databases, commonly used to benchmark point cloud quality metrics [29–32]. Crowdsourcing methodologies have also been studied [2,33] as a method of subjective evaluation.

2.3. Objective quality evaluation

Objective quality evaluation metrics are quite important for coding method developers, as subjective evaluation is time-consuming and typically requires careful planning and design.

Objective quality evaluation metrics can be divided into full-reference (the distorted data is compared with the original data), reduced-reference (features extracted from the distorted and original data are compared), and no-reference (only the distorted data is used for the quality estimation). Furthermore, in the case of point clouds, the quality evaluation metrics can also be divided into three categories: geometry only (only use the geometry information for the quality estimation), color only (only use the color attributes), and geometry plus color (use both geometry and color attributes). Furthermore, objective quality evaluation metrics can be classified as image-based or model-based metrics specifically developed for point cloud quality evaluation [34]. For instance, image-based metrics like PSNR and SSIM can be computed using a point cloud representative of 2D views.

The aforementioned study [6] evaluated a group of point-based metrics and concluded that point to point and point to plane metrics [16] using the mean squared error (MSE) were the best performing ones and provided a good representation of the subjective evaluation. Later, a benchmarking study using a 2D experimental setup for the subjective assessment was reported [35], which included a broader selection of objective quality metrics. The PCQM [36] and PointSSIM [37] showed the best performances in terms of correlation with the Mean Opinion Score (MOS). Moreover, the image metric Multiscale Structural Similarity index [38] (MS-SSIM) computed over the video generated for the 2-D visualization of the point cloud also revealed a good representation of the subjective evaluation.

(a) *Soldier* (b) *Guanyin* (c) *Longdress* (d) *Rhetorician* (e) *Romanoillamp* (f) *Bumbameuboi*

Fig. 1. Point Cloud testing set.

The metrics considered for this evaluation are the Point to Point (D1) [16], Point to Plane (D2) [16], Plane to Plane [39], the PSNR MSE YUV metric (Color PSNR of the luminance channel and each chroma component), the Point Cloud Structural Similarity (PointSSIM) metric [37], the Point Cloud Quality Metric (PCQM) [36], the Point to Distribution Metric (P2Dist) [40], the Reduced Reference Point Cloud Metric (PCM_{RR}) [41], the GraphSIM Metric [42] and the Color Histogram Metric [43]. Several no-reference metrics were also considered for this study, namely the IT-PCQA [44], MM-PCQA [45], ResSCNN [46], VQA-PC [47] and NR-3DQA [48]. While there are several other works available in the literature [49–52], the selected ones currently have a public implementation. A short revision of these metrics is presented in Section 3.4.

Some objective evaluation methods depend on the normals calculation. Taking this in consideration, two different computation methods were tested as in [53], namely the quadric fitting of Cloud Compare (CC), using a radius (R) of five, ten, and twenty [54], and the *K Nearest Neighbor* (KNN) with K of six, ten, and eighteen [55] using MeshLab (ML) plane fitting. Furthermore, image metrics were also considered, namely the FSIM and FSIMc [56], SSIM [57] and MSSIM [38], VIFp [58] and PSNR metrics.

3. Quality evaluation study

The methodology reported by Perry et al. [6] was selected for this study, as it has been proven to be very effective on the subjective evaluation of point clouds. To summarize, the subjects are shown a set of videos showing both the reference and distorted point clouds. Prior to the evaluation, the subjects visualized a small sequence of videos depicting a point cloud not included in the evaluation. The collected mean opinion scores (MOS) are then aggregated and compared to the predicted MOS values by the objective quality metrics. General information regarding the subjective evaluation is found in Section 3.3, while Section 3.5 reports the results of the objective evaluation, presenting both global and individual correlations and the study on the normal vectors influence.

3.1. Selected data

Fig. 1 represents the set of six point clouds containing geometry and texture information used in this study. The set consists of frame 0690 of the *Soldier* and frame 1300 of the *Longdress* dynamic point clouds representing human figures,[2] *Romanoillamp* and *Bumbameuboi* point clouds,[3] and *Rhetorician* and *Guanyin* point clouds, from the EPFL dataset,[4] representing cultural heritage artifacts. Moreover, frame 1550 of the *Redandblack* point cloud[5] was selected for the training session conducted prior to the subjective evaluations. This set includes only point clouds representing objects, as this was the main target of the tested codecs.

Table 1
Point cloud characteristics.

Point cloud	Sparcity (K=20)	Color gamut volume	Y deviation	Cb deviation	Cr deviation
Soldier	1.726	0.2%	0.095	0.009	0.009
Guanyin	1.748	1.3%	0.144	0.025	0.027
Longdress	1.730	4.6%	0.114	0.060	0.065
Rhetorician	1.764	0.4%	0.120	0.026	0.021
Romanoillamp	2.204	1.5%	0.094	0.020	0.012
Bumbameuboi	6.661	4%	0.132	0.070	0.067

Table 2
Parameters for the V-PCC codec.

V-PCC					
Rate	R01	R02	R03	R04	R05
Geometry QP	36	32	28	20	16
Texture QP	47	42	37	27	22
Occupancy Map			4		2

Table 3
Parameters for the G-PCC codec.

G-PCC					
Rate	R01	R02	R03	R04	R05
QP	46	40	34	28	22
pQS[a]	0.25	0.5	0.75	0.857	0.9375

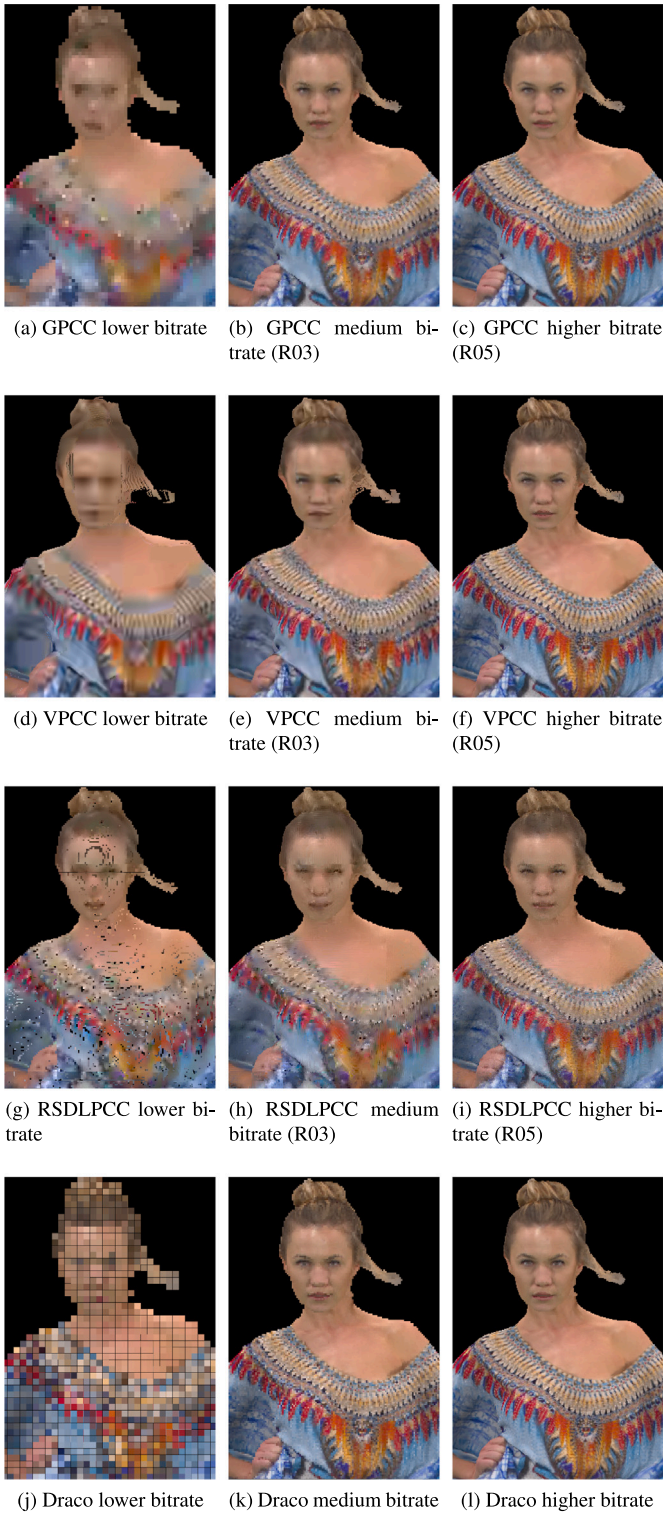
[a] pQS values were wrongly reported in EI2022 [1].

Table 1 depicts point cloud characteristics, namely the point cloud sparsity, color gamut volume, and standard deviation of each color channel of the YCbCr color space. The sparsity is defined as the average distance between each point and its 20 nearest neighbors, averaged over the entire point cloud. The color gamut volume is defined as the volume of the convex hull of the distribution of color points in the YCbCr color space. The test set reveals a small variation in sparsity. The sparsest point cloud is *Bumbameuboi*, followed by *Romanoillamp*. The rest of the point clouds in the set show very similar sparsity values. The set is a bit more diverse in terms of color gamut volume. The *Longdress* and *Bumbameuboi* point clouds present the highest values.

3.2. Dataset encoding

The dataset was encoded with the codecs under consideration. Tables 2 and 3 show the parameters used to encode the dataset with V-PCC and G-PCC, respectively, while Table 4 shows the settings for the Draco codec. The encoding results for RS-DLPCC were kindly provided by the authors [8]. As with any other deep learning based codec, the different bit-rates are obtained with a different configuration that results from training with a different rate/distortion tradeoff optimization.

[2] https://jpeg.org/plenodb
[3] http://uspaulopc.di.ubi.pt
[4] https://www.epfl.ch/labs/mmspg/downloads/pointxr/
[5] https://jpeg.org/plenodb



(a) GPCC lower bitrate (b) GPCC medium bitrate (R03) (c) GPCC higher bitrate (R05)

(d) VPCC lower bitrate (e) VPCC medium bitrate (R03) (f) VPCC higher bitrate (R05)

(g) RSDLPCC lower bitrate (h) RSDLPCC medium bitrate (R03) (i) RSDLPCC higher bitrate (R05)

(j) Draco lower bitrate (k) Draco medium bitrate (l) Draco higher bitrate

Fig. 2. Crop area of decoded results for *Longdress*.

Table 4
QP values for the Draco codec.

Rate	R01	R03	R05
QP	7	9	10

After encoding the test set, both V-PCC and G-PCC codecs had five quality levels, while RS-DLPCC had four quality levels and Draco had three quality levels. Taking the references into account, a total of 108 scores were obtained for each subject. Some examples of decompressed results for the *Longdress* point cloud are represented in Fig. 2, where typical distortions caused by each codec can be observed, such as the reduction in resolution typical of octree-based codecs or Draco, the typical smoothing that can be found on projection-based codecs, and the appearance of empty spaces typical of machine learning-based codecs.

3.3. Subjective experiment

In this section, the quality study of EI2022 [1] is presented. The subjective quality evaluation was conducted at the Multimedia Quality Laboratory of the Universidade da Beira Interior. In this section, a description of the subjective quality evaluation is reported, detailing the visualization procedure (Section 3.3.1) and the evaluation procedure (Section 3.3.2).

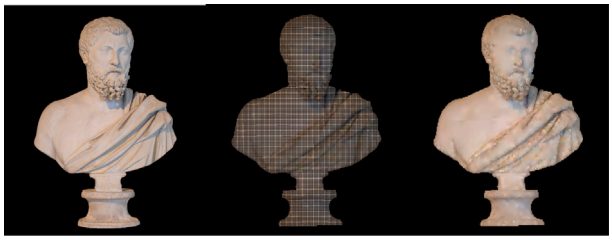
3.3.1. Visualization of point cloud content

The reference and distorted point clouds were rotated in a video side by side in a subjective double stimulus evaluation. For all point clouds, a complete rotation over the vertical axis was applied. At each rotation degree, an image representing the point cloud view was extracted, obtained using the Point Cloud Library (PCL)[6] Visualizer mode. Videos were created using the FFMPEG software. To ensure no compression was applied to the extracted frames, the stream copy option in FFMPEG was used.[7] The point cloud views were rendered at 30 fps. Each view represented a 1 degree rotation, resulting in 12 s videos displayed with a 1920 × 1080 resolution.

In some cases, the point size was manipulated to provide an improved visual representation. If the surface of the point cloud has some transparency or empty spaces, the subjects will see the opposite (or inner) part of the point cloud, creating a very bad quality perception [22,25]. This manipulation is important to avoid this perceptual effect by creating continuous surfaces.

However, it must be ensured that this manipulation does not mask compression artifacts. An example is represented in Fig. 3, namely for the *Rhetorician* point cloud, for the first rate of G-PCC. It can be observed that without manipulating the point size (Fig. 3(b)), parts of the point cloud are missing, and the opposite part of the point cloud is visible. This tends to happen when encoding at low rates, as the points are usually rather far from each other. When the point size is increased (Fig. 3(c)), the artifacts created by the codec are still noticeable, but the inner part of the point cloud is not visible anymore. It should be noted that each point cloud present in both the subjective evaluation and training session must be analyzed and the point size manipulated individually, as different rates and different types of artifacts require different point sizes. The point size for each decoded point cloud used in the subjective evaluation is described in Table 5, for the reference point cloud and for all different codec rates. In most cases, only the lower rates require manipulation, while the reference point clouds do not need any at all. Even so, there are cases where the reference point cloud needs to be manipulated, namely for the *Romanoillamp* and *Bumbameuboi* point clouds. As shown in Table 1, these are the sparsest point clouds in the dataset. Since this manipulation is closely related to the distance between points, the point size of the reference point cloud was manipulated as well.

[6] https://pointclouds.org
[7] https://ffmpeg.org/ffmpeg.html



(a) Reference: Point size 1 (b) G-PCC R01: Point Size 1 (c) G-PCC R01: Point Size 6

Fig. 3. Point size example for the *Rhetorician* point cloud.

Table 5
Selected point size for each content.

Content	Reference	V-PCC					G-PCC				
		R01	R02	R03	R04	R05	R01	R02	R03	R04	R05
Soldier	1	1	1	1	1	1	6	4	1	1	1
Guanyin	1	1	1	1	1	1	6	4	1	1	1
Longdress	1	1	1	1	1	1	6	4	1	1	1
Rhetorician	1	1	1	1	1	1	6	4	1	1	1
Romanoillamp	2	2	2	2	2	2	3	2	2	2	2
Bumbameuboi	4	4	4	4	4	4	6	4	4	4	4
		RS-DLPCC					Draco				
Soldier	1	–	6	5	1	1	6	–	2	–	1
Guanyin	1	–	6	4	1	1	6	–	2	–	1
Longdress	1	–	6	4	1	1	6	–	2	–	1
Rhetorician[a]	1	–	6	4	1	1	8	–	3	–	2
Romanoillamp[a]	2	–	7	3	2	2	6	–	2	–	2
Bumbameuboi	4	–	20	9	8	7	6	–	4	–	4

[a] Point sizes for Draco were wrongly reported in EI2022 [1].

Table 6
Subject Information.

Males	Females	Overall	Age span	Average age
10	6	16	21–33	24.75

3.3.2. Experimental procedure

Prior to the evaluation, the subjects were shown a sequence of four videos depicting the chosen point cloud for the training session (*Redandblack*), representing four different levels of degradation, so that they could familiarize themselves with the distortion artifacts created by the codecs. For the evaluation itself, the Double Stimulus Impairment Scale was selected. The subjects were shown both the reference and the distorted video representations of the point clouds, rotating side by side. Then the subjects were asked to evaluate the quality of the distorted point cloud on a five-level rating scale (1: very annoying, 2: slightly annoying, 3: annoying, 4: perceptible but not annoying, 5: imperceptible).

Distorted versions of the same content were never visualized one after the other. To avoid biases, half of the subjects were shown videos with the reference on the right and the distorted content on the left, and vice-versa. Additionally, hidden reference–reference pairs were included in the test sequence. All sessions were conducted at the Multimedia Quality Laboratory Group of Universidade da Beira Interior, using a 47-inch, FULL HD LG 47LA860V, with the evaluation environment following the specifications in [59]. Table 6 contains information on the subjects that performed the subjective evaluation.

After the subjective evaluation, all the scores were aggregated. Then the MOS and the 95% confidence intervals (CIs), assuming a Gaussian distribution, were computed for each content. The bitrate, measured in bits per point (bpp), is computed by taking the number of bits of the encoded content and dividing it by the number of points of the original content. Fig. 4 shows the MOS results. V-PCC provides the best quality scores, followed by G-PCC and the RS-DLPCC. Although Draco achieves very high scores, they are obtained at extremely high bitrates, making it the worst performing codec in this evaluation. The point cloud *Bumbameuboi* shows a different behavior from the other point clouds. This point cloud is rather sparse (as shown in Table 1) when compared with the others, which influences the codecs performance. This reveals that further studies need to be conducted in the future to address sparse point clouds, as they are commonly generated by multiple applications and with common acquisition technologies, such as LiDAR.

3.4. Description of the objective quality metrics

The quality metrics considered for this evaluation are classified as follows:

- Full-reference geometry only (FR, GEO): Point to Point (PSNR D1), Point to Plane (PSNR D2) and Plane to Plane (PL2Plane) and Point cloud Structural Similarity (PointSSIM) normal-based features;
- Full-reference color only (FR, COL): Point to attribute (PSNR YUV), PointSSIM luminance-based features and Color Histogram Metric [43];
- Full-reference geometry and color ($FR, GEO + COL$): Point Cloud Quality Metric (PCQM), Point 2 Distribution and GraphSIM;
- Reduced-reference color and geometry (RR, GEO+COL): Reduced reference point cloud metric (PCM_{RR});
- No-Reference geometry and color (NR, $GEO + COL$): IT PCQA, MM-PCQA, ResSCNN, NR 3DQA and VQA PC.

The PSNR D1 [16] metric computes the mean value of the geometric distance between the corresponding points in the reference and the distorted point cloud. The PSNR D2 [16] metric is computed by measuring the geometric distance between the distorted point cloud and the fitted reference point cloud local planes. For PSNR D1 and PSNR D2 metrics, the MSE or Hausdorff distances, are typically used as a measure of the geometric error. The geometry PSNR ratio is computed for both the MSE and Hausdorff distances using:

$$\text{PSNR} = 10\log_{10}\frac{3\times(2^p-1)^2}{\max\left\{\epsilon_{R,D},\epsilon_{D,R}\right\}}$$

Where $\epsilon_{R,D}$ represents the distance (MSE or Hausdorff) between the reference and distorted point clouds and $\epsilon_{D,R}$ between the distorted and the reference, while p represents the geometric resolution in number of bits.

For quality estimation, the PL2PLane metric [39] considers the angular differences between the point cloud normal vectors. The estimated plane angle between the distorted and reference point clouds is computed. The final metric is defined as a weighted mean using pooling estimators based on the mean squared (RMS) or mean squared error (MSE). Only the estimator that provided the best results is shown.

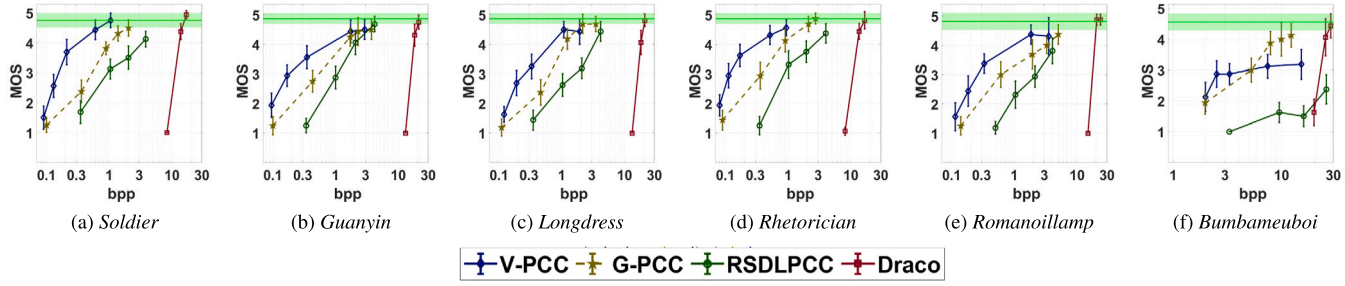
The Point 2 distribution [40] metric considers the *Mahalanobis* distance to assess geometry, color, or jointly geometry and color distortions. In this study only the joint geometry and color results were considered.

PCQM [36] uses a weighted linear combination of curvature and color information measures to predict the visual quality of a distorted point cloud.

GraphSIM [42] extracts geometric keypoints from the point cloud and then uses graph similarity to evaluate the distortions in the point clouds.

The PointSSIM metric [37] computes the similarity between geometry, normal vector, luminance, and curvature features.

For the feature extraction, dispersion statistics are computed using one of the available estimators, namely the median (m), variance (σ^2), mean absolute deviation (μ_{AD}), median absolute deviation (m_{AD}),



(a) *Soldier* (b) *Guanyin* (c) *Longdress* (d) *Rhetorician* (e) *Romanoillamp* (f) *Bumbameuboi*

Fig. 4. MOS vs bpp with 95% confidence intervals. The green bar in top represents the confidence interval obtained for the reference point cloud.

coefficient of variation (COV), and quartile coefficient of dispersion (QCD). The estimators will be applied over a number of K nearest neighbors, set to 12 [37]. Three different pooling methods are also considered, namely the arithmetic mean (mean), MSE, and RMS. In this study, the luminance features were considered, as they are the ones that usually provide the best results [37], as well as normal-based features, in order to analyze the performance of the considered normal computation methods. It should be noted that, when considering normal-based features, the metric does not consider color information. Finally, all the pooling methods and all the dispersion statistics were considered, and only the best performing ones are shown for each of the selected features.

The color Histogram [43] metric extracts color features from the reference and distorted point clouds and compares the resulting color histogram for each point cloud.

The PSNR YUV metric [60] computes the error of the color values between the identified point in the reference and the distorted point cloud. The identification process is conducted using the nearest neighbor algorithm, and an individual error for each color channel is computed for the identified points based on the Euclidean distance. The overall PSNR YUV is computed by weighting each color channel as $Y : U : V = 6 : 1 : 1$, using:

$$\text{PSNR} = 10 \log_{10} \frac{p^2}{\epsilon}$$

Where p is set to 255, because the considered point clouds have a color depth of 8 bit, and ϵ represents the MSE or Hausdorff distances.

The PCM_{RR} [41] metric extracts a small set of geometry, color, and normal features that are used to predict the visual quality of the content under evaluation. Then, it uses a linear optimization algorithm in order to find the best weights for each extracted feature.

The no-reference metrics are mostly based on deep learning. The ResSCNN [31] metric is based on hierarchical feature extraction using sparse convolution blocks with residual connections. VQA PC [61] generates three different video sequences, with different point cloud rotations. Those sequences contain multi-frame temporal information. The ResNet 3D is modified as the feature extraction methodology, learning the correlation between features of the video sequences and the MOS. NR-3DPCA [62] considers an SVR trained with 3D natural scene statistics and entropy features from the geometric and texture domains to evaluate the quality of both point clouds and meshes. IT-PCQA [44] is based on 3D to 2D projections and using unsupervised adversarial domain adaptation. MM-PCQA [45] renders the point clouds into 2D images after splitting them into sub-models. Those sub-models represent local geometric distortions. The images are encoded with point based and image-based neural networks. The final score is obtained using symmetric cross-modal attention. The default implementation of the no-reference metrics was used. The authors of NR-3DQA train their metrics using the Waterloo [29] database. The developers of IT PCQA and MM PCQA are trained using the SJTU-PCQA [30] database. Finally, ResSCNN is trained using the LS-PCQA [31] database.

The considered image metrics were computed for each individual frame of the tested videos. The final metric value was the average of the individual values for all the frames of each video (360).

3.5. Performance of objective point cloud quality metrics

The subjective evaluation (EI2022) results are used to validate the considered point cloud objective quality metrics. Furthermore, the BASICS validation dataset [2] is also considered. The validation subset of this publicly available database contains 15 point clouds, coded with V-PCC [4], G-PCC predlift [4], G-PCC RAHT [4], and GeoCNN [12]. This results in 300 distorted point clouds. The subjective evaluation protocol used a double-sided stimulus, and the subjective scores were obtained with crowdsourcing. For the considered metrics, the statistical measures proposed in [63] were computed, specifically the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Root-Mean Squared Error (RMSE) and the Outlier Ratio (OR). The predicted MOS for each of the objective metrics was computed by applying a logistic fitting function to the objective scores, as specified in [63] and usually done for objective metrics benchmarking [64].

Table 7 reports the performance of the different metrics for both datasets. Objective quality metrics that only consider luminance information are noted as LUM. Note that for the metrics dependent on normal calculations, only the best of the two considered methods is shown. Table 7 also reports the performance of the metrics for the BASICS validation dataset. For the metrics that depend on the normals, the best result of the EI2022 dataset is reported, as the influence of the different normal computation methods will be studied further ahead (Section 4.3). It can be observed that PCQM is the best performing metric for both datasets, followed by GraphSIM. Furthermore, the conclusions from both datasets are quite similar. The Hausdorff distances show very low performance compared to their MSE counterparts. The color histogram metric and the PCM_{RR} show a significant performance decrease. For both datasets, the no-reference metrics failed to provide accurate representations.

Regarding the image-based metrics, the VIFp reveals higher correlation values for both datasets, but its performance is lower than the best performing point cloud metrics.

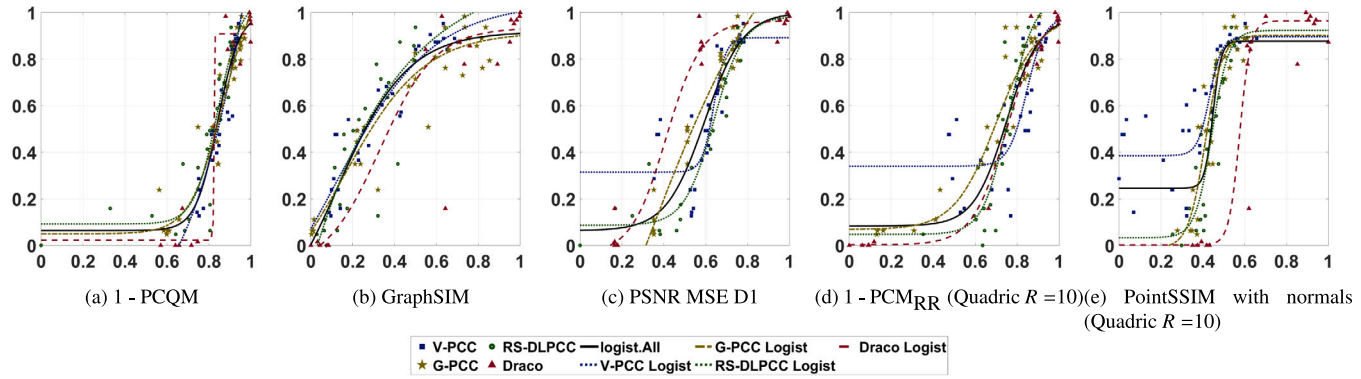
Fig. 5 shows the metric vs. MOS plots for the best five metrics and the logistic curves. The MOS values were normalized between 0 and 1, using a min–max normalization, as recommended by ITUT-BT500 [59]. Those logistic functions were used for the computation of the predicted MOS values from the metric values. Furthermore, as PCQM and PCM_{RR} values decrease with the increase in quality, their results are represented as 1-PCQM and 1-PCM_{RR} to allow an easier comparison with the other metrics.

Table 8 shows the metric correlation for each codec for the EI2022 dataset. Draco and G-PCC show the highest correlation values, with PCC and $SROCC$ values all above 0.93 for the full-reference metrics and for the reduced-reference metric PCM_{RR}.

Tables 9–11 show the influence of the normal computation method for metrics depending on them. The normals computed with the Cloud Compare quadric fitting algorithm present the best results, always surpassing Meshlab K nearest neighbors. Table 12 shows the variation of the PointSSIM metric regarding the estimators. For normal based

Table 7
Metrics Performance (best results in bold, and second best results in italic).

Metric	R (Quadric)	Type	Features	EI2022 [1]				BASICS [2]			
				PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
PSNR MSE D1 [16]	–	FR	GEO	0.890	0.884	0.148	0.618	0.894	0.800	0.126	0.760
PSNR MSE D2 [16]	20	FR	GEO	0.851	0.847	0.169	0.608	0.923	*0.836*	0.108	0.693
PSNR HAU D1 [16]	–	FR	GEO	0.574	0.507	0.264	0.755	0.437	0.341	0.253	0.840
PSNR HAU D2 [16]	10	FR	GEO	0.603	0.616	0.257	0.784	0.444	0.359	0.252	0.820
PL2PLane [39]	20	FR	GEO	0.846	0.795	0.172	0.667	0.650	0.628	0.212	0.877
PointSSIM normal-based [37]	10	FR	GEO	0.869	0.867	0.160	0.588	0.828	0.727	0.158	0.790
PointSSIM Luma-based [37]	–	FR	LUM	0.832	0.808	0.179	0.686	0.718	0.677	0.194	0.840
PSNR MSE YUV [60]	–	FR	COL	0.725	0.737	0.222	0.698	0.638	0.567	0.215	0.907
PSNR HAU YUV [60]	–	FR	COL	0.471	0.447	0.283	0.833	0.248	0.285	0.271	0.940
Color Histogram Metric [43]	–	FR	COL	0.797	0.820	0.196	0.637	0.497	0.428	0.244	0.883
PCQM [36]	–	FR	GEO+LUM	**0.944**	**0.928**	**0.106**	**0.480**	**0.927**	**0.849**	**0.105**	**0.690**
Point 2 Distribution [40]	–	FR	GEO+COL	0.778	0.794	0.204	0.747	0.748	0.612	0.186	0.847
GraphSIM [42]	–	FR	GEO+COL	*0.907*	*0.893*	*0.137*	*0.500*	*0.924*	0.817	*0.108*	*0.663*
PCM_{RR} [41]	10	RR	GEO+COL	0.890	0.871	0.147	0.529	0.612	0.516	0.221	0.867
IT PCQA [44]	–	NR	GEO+COL	0.525	0.317	0.275	0.853	0.022	0.049	0.281	0.940
MM PCQA [45]	–	NR	GEO+COL	0.761	0.722	0.210	0.735	0.721	0.545	0.194	0.923
ResSCNN [46]	–	NR	GEO+COL	0.433	0.413	0.291	0.902	0.154	0.123	0.277	0.940
NR 3DQA [48]	–	NR	GEO+COL	0.236	0.245	0.315	0.843	0.152	0.110	0.281	0.940
VQA PC [47]	–	NR	GEO+COL	0.566	0.370	0.267	0.892	0.463	0.397	0.248	0.923
FSIM [56]		FR	Image	0.843	0.842	0.17	0.608	0.586	0.477	0.228	0.877
FSIMc [56]		FR	Image	0.844	0.843	0.173	0.608	0.586	0.476	0.228	0.877
SSIM [57]		FR	Image	0.860	0.861	0.163	0.667	0.599	0.506	0.225	0.887
MSSIM [38]		FR	Image	0.850	0.850	0.170	0.676	0.538	0.423	0.237	0.897
VIFp [58]		FR	Image	0.886	0.874	0.149	0.569	0.630	0.548	0.218	0.880
PSNR		FR	Image	0.793	0.815	0.197	0.706	0.344	0.267	0.264	0.930



(a) 1 - PCQM (b) GraphSIM (c) PSNR MSE D1 (d) 1 - PCM_{RR} (Quadric R =10) (e) PointSSIM with normals (Quadric R =10)

Fig. 5. Relation between metrics and MOS, and Logistic fitting curve for the EI 2022 database.

features, only the Cloud Compare quadric fitting is considered as it provides better results. The recommended settings for the metric were considered the baseline value (considering luminance values and the σ^2 as a dispersion statistic [37], and the mean was used as a pooling estimator, as set in the default implementation of the metric[8]). It can be observed that the recommended baseline value achieves lower performance than most of the other available options. The best performance is obtained when considering the normal-based features, using COV as an estimator and the mean as the pooling method, for the EI2022 [1] dataset, and QCD as an estimator and MSE as the pooling method for the BASICS validation dataset [2]. When considering luminance-based features, the best performance is achieved when considering mAD as an estimator and MSE as a pooling method for both databases.

3.6. Statistical significance analysis

To further complement the study, a significance analysis evaluation using Krasula method was conducted [65]. This method evaluates the performance of objective metrics in two different analysis, notably "Different vs Similar" and "Better vs Worse".

In the first analysis, the distorted point clouds are paired and split into two different categories (Different and Similar). The "Different" category represents pairs with statistically significant differences, and the "Similar" the ones without. For each pair, a one way ANOVA followed by a Turkeys honest significance difference test [66] is conducted, measuring the statistical significance of the differences. The Krasula method assumes that the absolute difference of predictions made by the metrics for different pairs should be larger than the similar pairs. To quantify the performance, the Receiving Operating Characteristic (ROC) analysis is used, and the performance is expressed as the Area Under ROC Curve (AUC).

The second analysis (Better vs Worse) considers only the pairs that have statistically significant differences, by evaluating the performance of the metrics by identifying the better pairs with statistically significant differences. The performance is evaluated as correct classification percentage (CCp), and AUC.

The "Different vs Similar" analysis revealed that there are 523 pairs classified as Similar and, 4628 pairs classified as Different. From the Different pairs, 2124 were classified as better, and 2504 were classified as worse by the "Better vs Worse" analysis. Finally, the Krasula

[8] https://github.com/mmspg/pointssim

Table 8
Metrics correlation with the subjective scores for each codec, for the EI2022 [1] dataset.

Metric	PCC	SROCC	RMSE	OR
V-PCC				
PSNR MSE D1 [16]	0.894	0.861	0.141	0.367
PSNR MSE D2 (Quadric $R = 20$) [16]	**0.977**	**0.890**	**0.072**	**0.100**
PCM_{RR} (Quadric $R = 10$) [41]	0.827	0.805	0.178	0.600
PCQM [36]	0.949	0.943	0.099	0.200
PL2Plane (Quadric $R = 20$) [39]	0.883	0.845	0.149	0.400
PointSSIM luma-based [37]	0.915	0.896	0.126	0.333
PointSSIM normal-based (Quadric R=10) [37]	0.968	0.905	0.078	0.133
Color Histogram Metric [43]	0.919	0.910	0.103	0.400
GraphSIM [42]	0.934	0.919	0.093	0.233
IT PCQA [44]	0.001	0.061	0.260	0.700
MM PCQA [45]	0.810	0.746	0.153	0.567
ResSCNN [46]	0.569	0.592	0.214	0.633
NR 3DQA [48]	0.603	0.576	0.210	0.633
VQA PC [47]	0.150	0.096	0.272	0.733
G-PCC				
PSNR MSE D1 [16]	**0.978**	**0.890**	**0.070**	**0.033**
PSNR MSE D2 (Quadric $R = 20$) [16]	0.967	0.952	0.081	0.133
PCM_{RR} (Quadric $R = 10$) [41]	0.936	0.794	0.117	0.233
PCQM [36]	0.956	0.921	0.098	0.167
PL2Plane (Quadric $R = 20$) [39]	0.935	0.655	0.119	0.267
PointSSIM luma-based [37]	0.965	0.947	0.079	0.167
PointSSIM normal-based (Quadric $R = 10$) [37]	0.965	0.947	0.079	0.167
Color Histogram Metric [43]	0.930	0.924	0.116	0.233
GraphSIM [42]	0.949	0.852	0.099	0.200
IT PCQA [44]	0.514	0.454	0.270	0.667
MM PCQA [45]	0.785	0.755	0.197	0.433
ResSCNN [46]	0.413	0.491	0.285	0.733
NR 3DQA [48]	0.408	0.359	0.297	0.433
VQA PC [47]	0.536	0.443	0.270	0.567
RS-DLPCC				
PSNR MSE D1 [16]	**0.957**	**0.819**	**0.094**	**0.083**
PSNR MSE D2 (Quadric $R = 20$) [16]	0.938	0.893	0.111	0.250
PCM_{RR} (Quadric $R = 10$) [41]	0.831	0.822	0.179	0.375
PCQM [36]	0.897	0.885	0.142	0.333
PL2Plane (Quadric $R = 20$) [39]	0.948	0.913	0.102	0.167
PointSSIM luma-based [37]	0.832	0.808	0.179	0.686
PointSSIM normal-based (Quadric $R = 10$) [37]	0.884	0.820	0.122	0.500
Color Histogram Metric [43]	0.801	0.805	0.180	0.500
GraphSIM [42]	0.840	0.858	0.164	0.333
IT PCQA [44]	0.376	0.205	0.283	0.750
MM PCQA [45]	0.911	0.885	0.124	0.250
ResSCNN [46]	0.633	0.618	0.234	0.583
NR 3DQA [48]	0.037	0.016	0.302	0.750
VQA PC [47]	0.521	0.344	0.257	0.542
Draco				
PSNR MSE D1 [16]	0.990	0.833	0.062	0
PSNR MSE D2 (Quadric $R = 20$) [16]	0.990	0.809	0.063	0
PCM_{RR} (Quadric $R = 10$) [41]	0.993	0.853	0.051	0
PCQM [36]	0.980	0.787	0.087	0.056
PL2Plane (Quadric $R = 20$) [39]	**0.993**	**0.877**	**0.050**	**0**
PointSSIM luma-based [37]	0.927	0.820	0.169	0.056
PointSSIM normal-based(Quadric $R = 10$) [37]	0.993	0.839	0.053	0
Color Histogram Metric [43]	0.990	0.824	0.063	0
GraphSIM [42]	0.913	0.827	0.180	0.056
IT PCQA [44]	0.855	0.698	0.229	0.056
MM PCQA [45]	0.774	0.805	0.300	0.167
ResSCNN [46]	0.766	0.664	0.286	0.111
NR 3DQA [48]	0.649	0.535	0.339	0.167
VQA PC [47]	0.985	0.729	0.074	0.000

method also performs statistical significance tests. Notably, it employs the Hanley–Macneil Method [67], to compare AUC values from ROC analysis. Furthermore, the Fisher's [68] exact test is conducted to compare the percentage of correct recognition of the stimulus of higher quality.

4. Results discussion

Reliable metrics should represent similar behavior when compared with the subjective evaluation results. Comparing the plots of Fig. 4 with the plots of Figs. 6 to 8, it can be observed that most of the metrics achieve a similar representation of the subjective results.

Table 9
Influence of the plane estimation method on PSNR MSE D2 [16] metric.

Method	EI2022 [1]				BASICS [2]			
	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
Quadric $R = 5$	0.828	0.827	0.181	0.618	0.919	0.836	0.110	0.710
Quadric $R = 10$	0.834	0.831	0.179	0.608	0.920	**0.838**	0.109	0.723
Quadric $R = 20$	**0.851**	**0.847**	**0.169**	0.608	**0.923**	0.836	**0.108**	**0.693**
KNN $K = 6$	0.806	0.792	0.192	0.618	0.919	**0.838**	0.110	0.710
KNN $K = 10$	0.810	0.805	0.190	**0.598**	0.920	**0.838**	0.110	0.717
KNN $K = 18$	0.816	0.811	0.187	0.598	0.921	**0.838**	0.109	0.717

Table 10
Influence of the plane estimation method on PL2Plane [39] metric.

Method	EI2022 [1]				BASICS [2]			
	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
Quadric $R = 5$	0.806	0.792	0.191	0.735	**0.807**	**0.738**	**0.166**	**0.817**
Quadric $R = 10$	0.827	0.771	0.182	0.725	0.750	0.649	0.185	0.850
Quadric $R = 20$	**0.846**	**0.795**	**0.172**	**0.667**	0.650	0.628	0.212	0.877
KNN $K = 6$	0.318	0.297	0.309	0.892	0.453	0.457	0.250	0.880
KNN $K = 10$	0.387	0.357	0.301	0.863	0.451	0.502	0.250	0.863
KNN $K = 18$	0.473	0.373	0.286	0.833	0.434	0.493	0.253	0.893

Table 11
Influence of the plane estimation method on PCM_{RR} [41] metric.

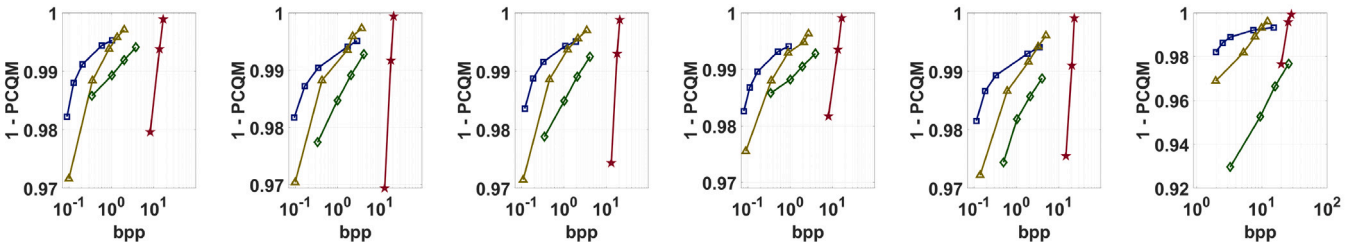
Method	EI2022 [1]				BASICS [2]			
	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
Quadric $R = 5$	0.884	0.872	0.151	0.588	**0.639**	**0.549**	**0.215**	**0.840**
Quadric $R = 10$	**0.890**	**0.871**	**0.147**	**0.529**	0.612	0.516	0.221	0.867
Quadric $R = 20$	0.842	0.843	0.174	0.559	0.594	0.493	0.224	0.883
KNN $K = 6$	0.670	0.653	0.242	0.784	0.608	0.576	0.221	0.857
KNN $K = 10$	0.673	0.642	0.240	0.765	0.617	0.581	0.219	0.847
KNN $K = 18$	0.781	0.767	0.201	0.667	0.616	0.565	0.219	0.857

4.1. Global analysis

The evaluation of the metrics representation is observed in the results of Table 7, revealing that the most commonly used metrics provide a somewhat accurate representation of the results when predicting the quality of point cloud compression in the presence of artifacts created by deep-learning technology, as most metrics achieved correlation values above 0.8. Table 7 also reveals that most of the full reference with geometry and color metrics achieve better results than the full reference with geometry only metrics. Table 7 further shows that most metrics achieve similar results for the BASICS validation database [2]. GraphSIM and PCQM are still the two best performing metrics. The main difference is that PSNR MSE D2 achieves a higher performance than PSNR MSE D1. The main cause for this is that the deep-learning solution GeoCNN [12] is optimized for the PSNR MSE D2 metric, which in turn increases the performance of the metric. Finally, Table 7 reveals that the no-reference metrics cannot accurately represent the quality of point cloud coding distortions. The metric that performed the best was MM-PCQA [45], achieving correlations above 0.7 for the subjective evaluation. For the BASICS validation dataset, the metric achieved a similar Pearson correlation, but the Spearman correlation value was very low. One reason for the low performance of the no-reference metrics is that they are trained on databases with a low number of point clouds that were distorted using the typical compression methods. The metrics are mainly trained on either the SJTU-PCQA or the Waterloo database, which contain several distortions that are not present in the currently used point cloud compression solutions. Fig. 5 also reveals that PSNR MSE D1 has a tendency to overestimate the quality of the deep learning solution, causing most of the samples to be below the logistic curve.

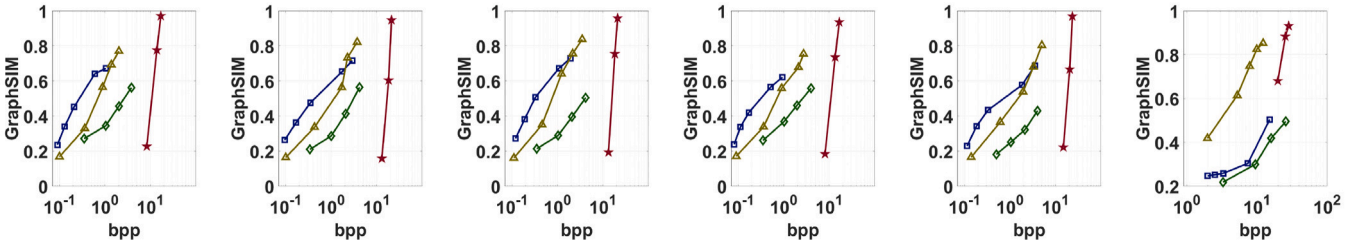
Table 12
Performance variation of PointSSIM [37] (Normals computed with the Cloud Compare quadric fitting method).

Features and pooling method	R	EI2022 [1]				BASICS [2]			
		PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
Using σ^2 as estimator									
Luma-based Mean (baseline)	–	0.798	0.789	0.195	0.706	0.718	0.677	0.194	0.840
Normal-based Mean	5	0.850	0.853	0.170	0.657	0.750	0.649	0.185	0.850
Normal-based Mean	10	0.863	0.863	0.164	0.588	0.856	0.742	0.145	0.740
Normal-based Mean	20	0.798	0.823	0.195	0.716	0.833	0.741	0.154	0.810
Using COV as estimator									
Luma-Based RMS	–	0.804	0.794	0.192	0.735	0.710	0.675	0.196	0.840
Normal-based Mean	5	0.866	0.862	0.162	0.578	0.840	0.730	0.152	0.720
Normal-bases Mean	10	**0.869**	**0.867**	**0.160**	**0.588**	0.853	0.732	0.147	0.777
Normal-based Mean	20	0.805	0.823	0.192	0.686	0.828	0.727	0.158	0.790
Using mAD as estimator									
Luma-based MSE	–	0.832	0.808	0.179	0.686	0.724	0.676	0.193	0.847
Normal-based MSE	5	0.819	0.830	0.187	0.637	0.843	0.735	0.151	0.717
Normal-based MSE	10	0.818	0.837	0.189	0.588	0.852	0.735	0.147	0.740
Normal-based MSE	20	0.778	0.804	0.204	0.735	0.827	0.715	0.158	0.783
Using QCD as estimator									
Luma-based MSE	–	0.828	0.811	0.181	0.647	0.734	0.687	0.190	0.830
Normal-based MSE	5	0.838	0.841	0.177	0.608	0.846	0.739	0.149	0.717
Normal-based MSE	10	0.826	0.843	0.185	0.618	**0.863**	**0.745**	**0.142**	**0.723**
Normal-based MSE	20	0.765	0.812	0.209	0.686	0.829	0.723	0.157	0.783
Using μAD as estimator									
Luma-based MSE	–	0.801	0.790	0.194	0.686	0.719	0.683	0.194	0.840
Normal-based Mean	5	0.844	0.844	0.173	0.637	0.835	0.725	0.154	0.733
Normal-based Mean	10	0.856	0.859	0.168	0.588	0.852	0.732	0.147	0.767
Normal-based Mean	20	0.795	0.822	0.197	0.696	0.824	0.722	0.159	0.807
Using SD as estimator									
Luma-based RMS	–	0.799	0.789	0.195	0.706	0.718	0.677	0.194	0.840
Normal-based Mean	5	0.853	0.858	0.169	0.627	0.836	0.725	0.154	0.720
Normal-based Mean	10	0.866	0.863	0.162	0.627	0.852	0.731	0.147	0.773
Normal-based Mean	20	0.804	0.819	0.192	0.716	0.827	0.727	0.158	0.787



(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Fig. 6. (1 - PCQM) vs BPP.

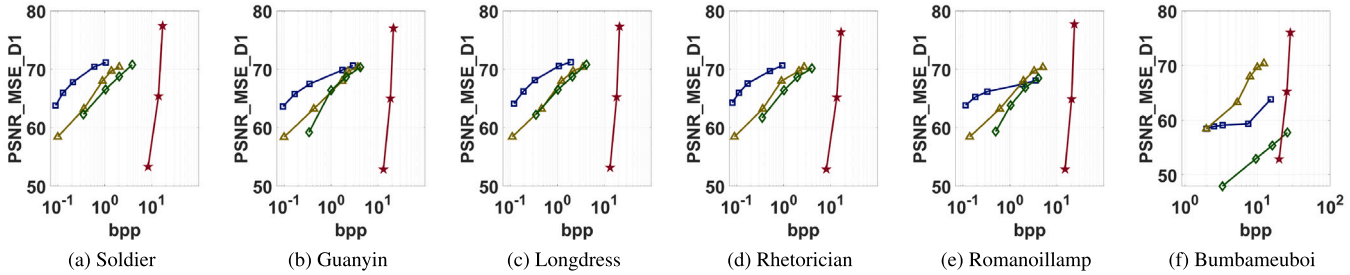


(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Fig. 7. GraphSIM vs BPP.

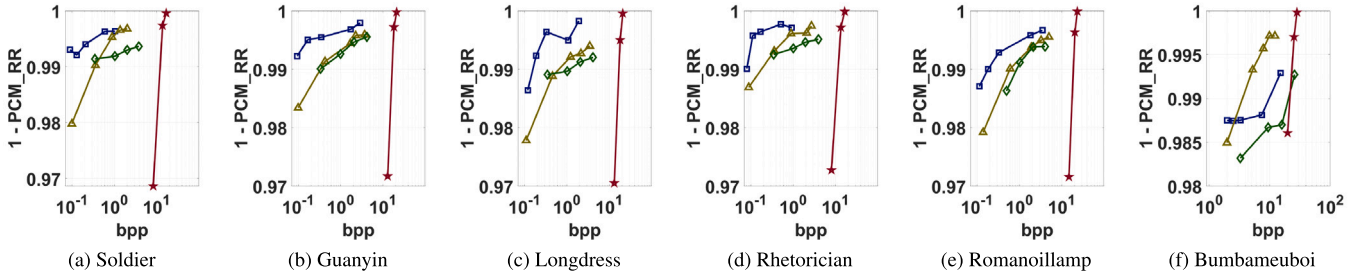
4.2. Individual codecs

Table 8 shows that G-PCC and Draco are the codecs that show higher correlation values for the different metrics. The distortions created by the codecs are octree/kd-tree based, that are simpler for the metrics evaluation. It is also observed that RS-DLPCC shows the worst correlation values, This codec tends to create blocking artifacts (Fig. 2), which are typically difficult to evaluate by the objective metrics [17].



(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Fig. 8. PSNR MSE D1 vs BPP.



(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Fig. 9. (1 - PCM_{RR}.) vs BPP. Normals Computed with Cloud Compare Quadric Fitting with $R = 10$.

Furthermore, this codec was the one that required the most point size adjustment, as shown in Table 5. Although this is a necessary procedure, it can create a gap between the subjective evaluation and the objective metrics. That effect might become more relevant for metrics that use color information. Another evidence is that objective metrics that only consider geometry tend to show similar values across codecs.

As a final remark, it can be observed that the full-reference geometry metrics have the best performance for each individual codec. However, they fail when they compare the different codecs between them. The joint geometry and color full reference metrics have lower performance, that might be caused by distortions introduced by G-PCC after encoding the geometry with RS-DLPCC. Furthermore, the full reference metrics have a great advantage over both the reduced-reference and no-reference methods.

The metrics that only consider geometry yield superior results than the metrics that only consider luminance or color. This observation reveals humans are very sensitive to the geometry of the point clouds, while being less demanding regarding the color information. Previous studies [69] conducted for 2D images, showed that humans are more sensitive to changes in the image structure than to color variations. These results seem to generalize this well know 2D fact to 3D content.

The joint quality metrics work around the balance between geometry and color, considering both types of information. PCQM is sensitive to the changes in curvature between the reference and distorted point clouds, which is revealed to be a very important feature for the definition of subjective quality. Moreover, luminance features are used in a weighted combination. GraphSIM processes the reference point cloud using a high-pass graph filter, which extracts structural information. This is used to extract keypoints that define neighborhoods where the distorted and reference point clouds luminance and chromatic information are compared. Hence, both metrics use structural information to predict the objective quality, along with color information.

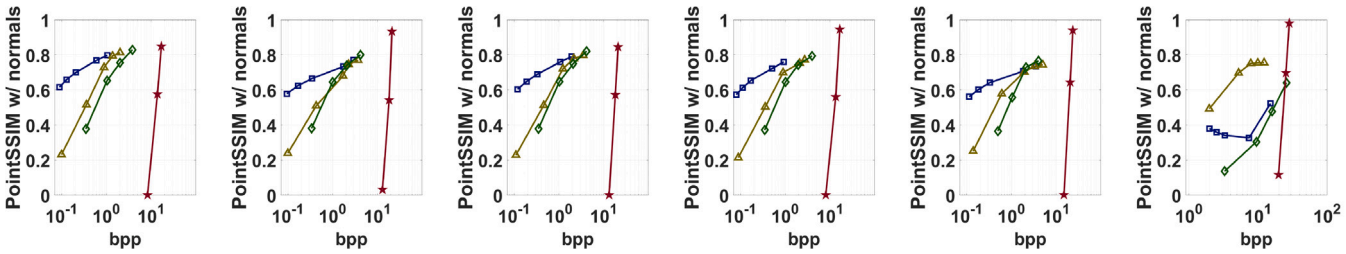
4.3. Normal plane influence

Regarding the influence of the methodology on the normal computations required for PSNR MSE D2, PL2Plane, PCM_{RR} and PointSSIM metrics, the conclusions are that the quadric fitting solution is the most suitable, with a radius of either 10 or 20, as those were the values that typically achieved the highest correlation scores. The PL2Plane and PCM_{RR} exhibits a better performance with a radius of 5 for the BASICS validation dataset. Both metrics have low performance with BASICS.

Plane estimation using the Meshlab *K Nearest Neighbors* failed to provide reliable normals for metrics computation, reaching values as low as 0.318 for *PCC*. On the contrary, Cloud Compare quadric fitting almost always achieved results that allowed reliable normal computation, as shown in Tables 9 to 12. In most cases, PCC and SROCC values above 0.8 were obtained. Consequentially, most correlations are slightly below 0.9. While those results are still acceptable, they reveal that a cautious approach is needed when using point cloud quality metrics for the evaluation of different compression technologies that are dependent on normals information.

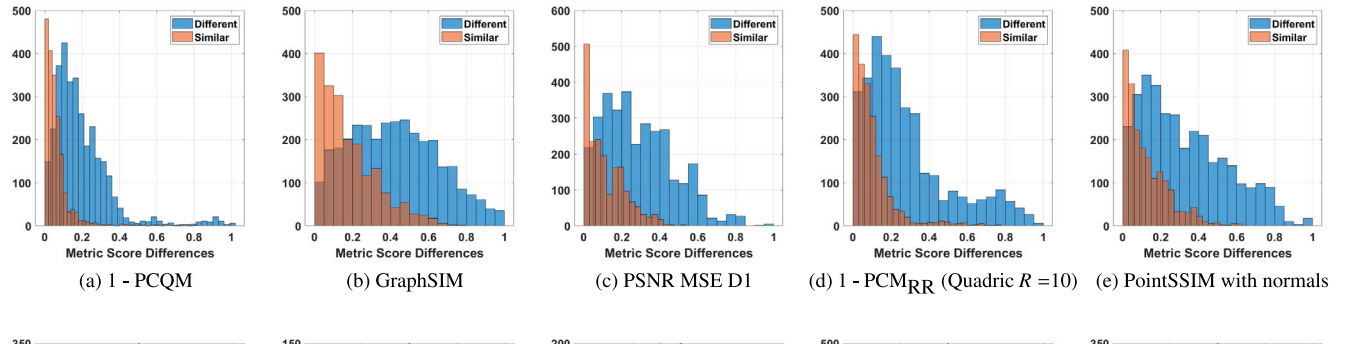
4.4. Rate distortion analysis

Figs. 6 to 10 show the metric vs. bpp (Bits Per Point) plots for the five best performing metrics for the EI2022 [1] dataset, namely PCQM [36], GraphSIM [42], MSE PSNR D1 [16], PCM_{RR} [41] and PointSSIM [37] with normal-based features. All metrics reveal occurrences where the quality swaps between the different encoders. In particular, the deep learning based solution (RS-DLPCC) tends to achieve better results in the objective evaluation than the ones obtained in the subjective evaluation. Most metrics also show very similar results between the G-PCC and the deep learning solution, even though the subjective evaluation revealed different quality levels. This shows that some caution is necessary when using these metrics to evaluate machine learning based solutions. The plots also reveal that both the PCQM (Fig. 6) and GraphSIM (Fig. 7) metrics show very similar behaviors for all coding solutions. Most metrics have monotonic growth with the bit rate, as is observed for the MOS. However, some exceptions can be observed, notably for PCM_{RR} (Fig. 9) and PointSSIM (Fig. 10), that in some cases have a non-monotonic behavior with the bitrate. It is important to emphasize that the subjective quality evaluation also has cases that do not reveal monotonic behavior. These cases happen when the balance between the geometry and the color quality is difficult to establish.



(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

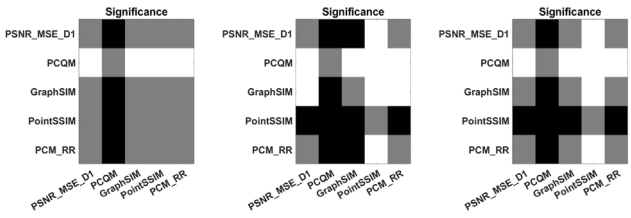
Fig. 10. PointSSIM (with normals) vs BPP. Normals Computed with Cloud Compare Quadric Fitting with $R = 10$, using the covariance (COV) dispersion statistic and the mean as a pooling estimator.



(a) 1 - PCQM (b) GraphSIM (c) PSNR MSE D1 (d) 1 - PCM_{RR} (Quadric R =10) (e) PointSSIM with normals

(f) 1 - PCQM (g) GraphSIM (h) PSNR MSE D1 (i) 1 - PCM_{RR} (Quadric R =10) (j) PointSSIM with normals (Quadric R =10

Fig. 11. The top row of plots depicts the metric scores differences for the "Different" and "Similar" analysis, for the EI2022 Dataset [1]. Each metric scores differences that are individually normalized within the minimum and maximum ranges. The height of the bars represents the number of occurrences. The bottom row shows the metric score differences for pairs categorized as "Better" and "Worse". Again, the metric score differences are individually normalized within the minimum and maximum ranges.



(a) Different vs Similar (AUC) (b) Better Worse (AUC) (c) Better Worse (CCp)

Fig. 12. Statistical Analysis results for the EI2022 Dataset [1]. Significance plots show that the performance of the method in the row is either significantly better (white), lower (black), or none of the previous (gray).

Table 13
Area Under ROC Curve (AUC) and Correct Classification Percentage (CCp) for the five best performing metrics.

Metrics	AUC Different vs Similar	AUC Better vs Worse	CCp
PCQM	0.8575	0.9954	0.9817
GraphSIM	0.8074	0.9885	0.9619
PSNR MSE D1	0.7965	0.9901	0.9489
PCM_{RR} normal-based ($R = 10$)	0.7986	0.9907	0.9508
PointSSIM normal-based ($R = 10$)	0.7884	0.9835	0.9402

4.5. Krasula method

Fig. 11 shows the results for each analysis for the five best performing metrics for the EI2022 [1] dataset. Table 13 shows the AUC values for both analysis and CCp for the referred metrics. The PCQM metric reveals the best results in both analysis. It achieves the highest AUC values and CCp. Furthermore, statistical significance tests were performed, to understand when a given metric has an increased performance over another, for both types of analysis. The results are shown in Fig. 12, where it can be observed that from the five best performing metrics, PCQM performs significantly better. It is interesting to observe that PSNR MSE D1 and GraphSIM show no statically significant differences. GraphSIM achieves a higher AUC in the Different vs Similar analysis,

while PSNR MSE D1 yields a slightly higher AUC in the Better vs Worse analysis.

5. Final comments

This study provides a benchmark for point cloud objective quality metrics in the presence of coding distortions. Several metrics provided a poor representation of the subjective quality for the different distortions caused by the considered coding methods. This makes it difficult to compare different coding technologies because they create different types of artifacts. It was observed that V-PCC, a projection-based codec that uses image coding technology, tends to be over-evaluated by the studied metrics for its low bit rates. Moreover, most metrics tend to over-evaluate the deep learning based solution, showing higher results than those observed in the subjective evaluation.

From the analyzed metrics, it was concluded that PCQM is the best performing metric for the evaluation of point cloud coding solutions. Furthermore, the Krasula test revealed that this metric shows statistically better results than the others. Finally, GraphSIM is the second best metric, achieving very good results for both datasets.

The study on the influence of the normal plane estimation on metrics that rely on them, revealed that the best method is cloud compare quadric fitting, as it always outperformed Meshlab KNN. From the results, it is recommended to use a radius of 20. Regarding PointSSIM, it was concluded that the best estimator for the luminance features is the mAD. For the normal features the best estimator is the COV closely followed by QCD. The normals should be computed using the quadric fitting with a radius of 10.

It is also advised to consider as much metrics as possible when evaluating point cloud coding solutions. A metric that never achieved the best correlation values may outperform the metric that is considered the best. However, when limited resources are available, it is recommended to always include PCQM and GraphSIM for point cloud coding solutions benchmarking, since they are the ones that consistently provided the most accurate representations of the subjective evaluations.

It was also observed that the MPEG coding solutions achieve the best performance. V-PCC reveals the highest performance, except for the *Bumbameuboi* point cloud. This result seems to be linked to the higher sparsity of this point cloud. Furthermore, the deep-learning solution shows promising results. It manages to have a very similar performance as G-PCC, although it has added scalability property. The Draco codec fails to provide state-of-the art results.

The point clouds used in this study are publicly available at https://github.com/JpcPrazeres/SPIC2024.

CRediT authorship contribution statement

Joao Prazeres: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Manuela Pereira:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Antonio M.G. Pinheiro:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was funded by FCT/MCTES, Portugal through national funds and when applicable co-funded EU funds under the projects UIDB/50008/2020 and PLive X-0017-LX-20.

References

[1] J. Prazeres, M. Pereira, A. Pinheiro, Quality analysis of point cloud coding solutions, Electron. Imaging 34 (17) (2022) 225–1–225–1.

[2] A. Ak, E. Zerman, M. Quach, A. Chetouani, A. Smolic, G. Valenzise, P. Le Callet, BASICS: Broad quality assessment of static point clouds in a compression scenario, IEEE Trans. Multimed. (2024) 1–13.

[3] R.B. Rusu, S. Cousins, 3D is here: Point cloud library (PCL), in: IEEE International Conference on Robotics and Automation, 2011, pp. 1–4.

[4] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, A. Tabatabai, An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC), APSIPA Trans. Signal Inf. Process. 9 (2020) e13.

[5] R.L. de Queiroz, P.A. Chou, Compression of 3D point clouds using a region-adaptive hierarchical transform, IEEE Trans. Image Process. 25 (8) (2016) 3947–3956.

[6] S. Perry, H.P. Cong, L.A. da Silva Cruz, J. Prazeres, M. Pereira, A. Pinheiro, E. Dumic, E. Alexiou, T. Ebrahimi, Quality evaluation of static point clouds encoded using MPEG codecs, in: 2020 IEEE International Conference on Image Processing, ICIP, 2020, pp. 3428–3432.

[7] A.F.R. Guarda, N.M.M. Rodrigues, F. Pereira, Point cloud coding: Adopting a deep learning-based approach, in: 2019 Picture Coding Symposium, PCS, 2019, pp. 1–5.

[8] A.F.R. Guarda, N.M.M. Rodrigues, F. Pereira, Deep learning-based point cloud geometry coding with resolution scalability, in: 2020 IEEE 22nd International Workshop on Multimedia Signal Processing, MMSP, 2020, pp. 1–6.

[9] J. Wang, H. Zhu, H. Liu, Z. Ma, Lossy point cloud geometry compression via end-to-end learning, IEEE Trans. Circuits Syst. Video Technol. 31 (12) (2021) 4909–4923.

[10] J. Wang, D. Ding, Z. Li, Z. Ma, Multiscale point cloud geometry compression, in: 2021 Data Compression Conference, DCC, 2021, pp. 73–82.

[11] M. Quach, G. Valenzise, F. Dufaux, Learning convolutional transforms for lossy point cloud geometry compression, in: 2019 IEEE International Conference on Image Processing, ICIP, 2019, pp. 4320–4324.

[12] M. Quach, G. Valenzise, F. Dufaux, Improved deep point cloud geometry compression, in: 2020 IEEE 22nd International Workshop on Multimedia Signal Processing, MMSP, 2020, pp. 1–6.

[13] A.F.R. Guarda, N.M.M. Rodrigues, F. Pereira, Adaptive deep learning-based point cloud geometry coding, IEEE J. Sel. Top. Sign. Proces. 15 (2) (2021) 415–430.

[14] J. Wang, D. Ding, Z. Li, X. Feng, C. Cao, Z. Ma, Sparse tensor-based multiscale representation for point cloud geometry compression, IEEE Trans. Pattern Anal. Mach. Intell. (2022) 1–18.

[15] J. Prazeres, R. Rodrigues, M. Pereira, A.M.G. Pinheiro, On the stability of point cloud machine learning based coding, in: 2022 10th European Workshop on Visual Information Processing, EUVIP, 2022, pp. 1–6.

[16] D. Tian, H. Ochimizu, C. Feng, R. Cohen, A. Vetro, Geometric distortion metrics for point cloud compression, in: 2017 IEEE International Conference on Image Processing, ICIP, 2017, pp. 3460–3464.

[17] J. Prazeres, R. Rodrigues, M. Pereira, A.M. Pinheiro, Quality evaluation of machine learning-based point cloud coding solutions, in: Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis, APCCPA '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 57–65.

[18] O. Devillers, P.-M. Gandoin, Geometric compression for interactive transmission, in: Proceedings Visualization 2000. VIS 2000 (Cat. No.00CH37145), 2000, pp. 319–326.

[19] C.-H. Wu, C.-F. Hsu, T.-K. Hung, C. Griwodz, W.T. Ooi, C.-H. Hsu, Quantitative comparison of point cloud compression algorithms with PCC arena, IEEE Trans. Multimed. 25 (2023) 3073–3088.

[20] C.-H. Wu, C.-F. Hsu, T.-C. Kuo, C. Griwodz, M. Riegler, G. Morin, C.-H. Hsu, PCC Arena: A benchmark platform for point cloud compression algorithms, in: Proceedings of the 12th ACM International Workshop on Immersive Mixed and Virtual Environment Systems, MMVE '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–6.

[21] E. Zerman, C. Ozcinar, P. Gao, A. Smolic, Textured mesh vs coloured point cloud: A subjective study for volumetric video compression, in: 2020 Twelfth International Conference on Quality of Multimedia Experience, QoMEX, 2020, pp. 1–6.
[22] 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, A. Skodras, Point cloud subjective evaluation methodology based on 2D rendering, in: 2018 Tenth International Conference on Quality of Multimedia Experience, QoMEX, 2018.
[23] E. Alexious, A.M.G. Pinheiro, C. Duarte, D. Matković, E. Dumić, L.A. da Silva Cruz, L.G. Dmitrović, M.V. Bernardo, M. Pereira, T. Ebrahimi, Point cloud subjective evaluation methodology based on reconstructed surfaces, in: Applications of Digital Image Processing XLI, SPIE, 2018.
[24] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, Subjective and objective quality evaluation of compressed point clouds, in: 2017 IEEE 19th International Workshop on Multimedia Signal Processing, MMSP, 2017, pp. 1–6.
[25] L.A. da Silva Cruz, E. Dumić, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, T. Ebrahimi, Point cloud quality evaluation: Towards a definition for test conditions, in: 2019 Eleventh International Conference on Quality of Multimedia Experience, QoMEX, 2019, pp. 1–6.
[26] E. Alexiou, I. Viola, T.M. Borges, T.A. Fonseca, R.L. De Queiroz, T. Ebrahimi, A comprehensive study of the rate-distortion performance in MPEG point cloud compression, APSIPA Trans. Signal Inf. Process. (2019).
[27] J. Prazeres, M. Pereira, A.M.G. Pinheiro, Subjective quality evaluation of point clouds with 3D stereoscopic visualization, in: 2022 IEEE International Conference on Image Processing, ICIP, 2022, pp. 2861–2865.
[28] S. Perry, L.A. Da Silva Cruz, J. Prazeres, A. Pinheiro, E. Dumic, D. Lazzarotto, T. Ebrahimi, Subjective and objective testing in support of the JPEG pleno point cloud compression activity, in: 2022 10th European Workshop on Visual Information Processing, EUVIP, 2022, pp. 1–6.
[29] Q. Liu, H. Su, Z. Duanmu, W. Liu, Z. Wang, Perceptual quality assessment of colored 3D point clouds, IEEE Trans. Vis. Comput. Graphics (2022) 1–1.
[30] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang, J. Sun, Predicting the perceptual quality of point cloud: A 3D-to-2D projection-based exploration, IEEE Trans. Multimed. 23 (2021) 3877–3891.
[31] Y. Liu, Q. Yang, Y. Xu, L. Yang, Point cloud quality assessment: Dataset construction and learning-based no-reference metric, ACM Trans. Multimedia Comput. Commun. Appl. 19 (2s) (2023).
[32] L. Hua, M. Yu, Z. He, R. Tu, G. Jiang, CPC-GSCT: Visual quality assessment for coloured point cloud based on geometric segmentation and colour transformation, IET Image Process. 16 (4) (2022) 1083–1095.
[33] S. Perry, L.A. Da Silva Cruz, E. Dumic, N.H. Thi Nguyen, A. Pinheiro, E. Alexiou, Comparison of remote subjective assessment strategies in the context of the JPEG Pleno Point Cloud Activity, in: 2021 IEEE 23rd International Workshop on Multimedia Signal Processing, MMSP, 2021.
[34] G. Lavoué, M.C. Larabi, L. Váša, On the efficiency of image metrics for evaluating the visual quality of 3D models, IEEE Trans. Vis. Comput. Graphics (2016).
[35] D. Lazzarotto, E. Alexiou, T. Ebrahimi, Benchmarking of objective quality metrics for point cloud compression, in: 2021 IEEE 23rd International Workshop on Multimedia Signal Processing, MMSP, 2021, pp. 1–6.
[36] G. Meynet, Y. Nehmé, J. Digne, G. Lavoué, PCQM: A full-reference quality metric for colored 3D point clouds, in: 2020 Twelfth International Conference on Quality of Multimedia Experience, QoMEX, 2020, pp. 1–6.
[37] E. Alexiou, T. Ebrahimi, Towards a point cloud structural similarity metric, in: 2020 IEEE International Conference on Multimedia & Expo Workshops, ICMEW, 2020, pp. 1–6.
[38] Z. Wang, E. Simoncelli, A. Bovik, Multiscale structural similarity for image quality assessment, in: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, vol. 2, 2003, pp. 1398–1402.
[39] E. Alexiou, T. Ebrahimi, Point cloud quality assessment metric based on angular similarity, in: 2018 IEEE International Conference on Multimedia and Expo, ICME, 2018, pp. 1–6.
[40] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, A point-to-distribution joint geometry and color metric for point cloud quality assessment, in: 2021 IEEE 23rd International Workshop on Multimedia Signal Processing, MMSP, 2021, pp. 1–6.
[41] I. Viola, P. Cesar, A reduced reference metric for visual quality evaluation of point cloud contents, IEEE Signal Process. Lett. 27 (2020) 1660–1664.
[42] Q. Yang, Z. Ma, Y. Xu, Z. Li, J. Sun, Inferring point cloud quality via graph similarity, IEEE Trans. Pattern Anal. Mach. Intell. 44 (6) (2022) 3015–3029.
[43] I. Viola, S. Subramanyam, P. Cesar, A color-based objective quality metric for point cloud contents, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6.
[44] Q. Yang, Y. Liu, S. Chen, Y. Xu, J. Sun, No-reference point cloud quality assessment via domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2022.
[45] Z. Zhang, W. Sun, X. Min, Q. Wang, J. He, Q. Zhou, G. Zhai, MM-PCQA: Multi-modal learning for no-reference point cloud quality assessment, in: E. Elkind (Ed.), Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23, International Joint Conferences on Artificial Intelligence Organization, 2023, pp. 1759–1767, Main Track.
[46] Y. Liu, Q. Yang, Y. Xu, L. Yang, Point cloud quality assessment: Dataset construction and learning-based no-reference metric, ACM Trans. Multimedia Comput. Commun. Appl. (2022).
[47] Z. Zhang, W. Sun, Y. Zhu, X. Min, W. Wu, Y. Chen, G. Zhai, Evaluating point cloud from moving camera videos: A no-reference metric, IEEE Trans. Multimed. (2023) 1–13.
[48] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, G. Zhai, No-reference quality assessment for 3D colored point cloud and mesh models, IEEE Trans. Circuits Syst. Video Technol. 32 (11) (2022) 7618–7631.
[49] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[50] M. Tliba, A. Chetouani, G. Valenzise, F. Dufaux, Point cloud quality assessment using cross-correlation of deep features, in: Proceedings of the 2nd Workshop on Quality of Experience in Visual Multimedia Applications, QoEVMA '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 63–68.
[51] M. Tliba, A. Chetouani, G. Valenzise, F. Dufaux, PCQA-Graphpoint: Efficient deep-based graph metric for point cloud quality assessment, in: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2023, pp. 1–5.
[52] R. Charles, H. Su, M. Kaichun, L.J. Guibas, PointNet: Deep learning on point sets for 3D classification and segmentation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE Computer Society, Los Alamitos, CA, USA, 2017, pp. 77–85.
[53] ISO/IEC JTC1/SC29/WG1M89044, JPEG pleno PC exploration study 4 results, 89th meeting, 2020.
[54] P. Bo, R. Ling, W. Wang, A revisit to fitting parametric surfaces to point clouds, Comput. Graph. 36 (5) (2012) 534–540, Shape Modeling International (SMI) Conference 2012.
[55] K. Klasing, D. Althoff, D. Wollherr, M. Buss, Comparison of surface normal estimation methods for range sensing applications, in: 2009 IEEE International Conference on Robotics and Automation, 2009, pp. 3206–3211.
[56] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[57] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[58] H. Sheikh, A. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.
[59] ITU-R BT.500-15, Methodologies for the subjective assessment of the quality of television images, 2023.
[60] E.M. Torlig, E. Alexiou, T.A. Fonseca, R.L. de Queiroz, T. Ebrahimi, A novel methodology for quality assessment of voxelized point clouds, in: A.G. Tescher (Ed.), in: Applications of Digital Image Processing XLI, vol. 10752, SPIE, International Society for Optics and Photonics, 2018, p. 107520I.
[61] Y. Fan, Z. Zhang, W. Sun, X. Min, N. Liu, Q. Zhou, J. He, Q. Wang, G. Zhai, A no-reference quality assessment metric for point cloud based on captured video sequences, in: 2022 IEEE 24th International Workshop on Multimedia Signal Processing, MMSP, 2022, pp. 1–5.
[62] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, G. Zhai, No-reference quality assessment for 3D colored point cloud and mesh models, IEEE Trans. Circuits Syst. Video Technol. 32 (11) (2022) 7618–7631.
[63] ITU-T P.1401, International telecommunication union, in: Methods, Metrics and Procedures for Statistical Evaluation, Qualification and Comparison of Objective Quality Prediction Models, 2012.
[64] P. Hanhart, M.V. Bernardo, M. Pereira, A.M. G. Pinheiro, T. Ebrahimi, Benchmarking of objective quality metrics for HDR image quality assessment, EURASIP J. Image Video Process. 2015 (2015) 1–18.
[65] L. Krasula, K. Fliegel, P. Le Callet, M. Klíma, On the accuracy of objective image and video quality models: New methodology for performance evaluation, in: 2016 Eighth International Conference on Quality of Multimedia Experience, QoMEX, 2016, pp. 1–6.
[66] J.W. Tukey, Comparing individual means in the analysis of variance, Biometrics 5 (2) (1949) 99–114.
[67] J.A. Hanley, B.J. McNeil, A method of comparing the areas under receiver operating characteristic curves derived from the same cases, Radiology 148 (3) (1983) 839–843.
[68] R.A. Fisher, On the interpretation of x^2 from contingency tables, and the calculation of P, J. R. Stat. Soc. 85 (1) (1922) 87–94.
[69] M.V. Bernardo, A.M.G. Pinheiro, P.T. Fiadeiro, M. Pereira, Image quality under chromatic impairments, ACM Trans. Appl. Percept. 14 (1) (2016).

5.2 Performance Analysis of Deep Learning-based Lossy Point Cloud Geometry Compression Coding Solutions

J. Prazeres, R. Rodrigues, M. Pereira and A. M. G. Pinheiro,
”Performance Analysis of Deep Learning-Based Lossy Point Cloud Geometry Compression Coding Solutions,”
in IEEE Access, vol. 13, pp. 76000-76015, 2025,
doi: 10.1109/ACCESS.2025.3561895

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 February 2025, accepted 6 April 2025, date of publication 17 April 2025, date of current version 6 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3561895

RESEARCH ARTICLE

Performance Analysis of Deep Learning-Based Lossy Point Cloud Geometry Compression Coding Solutions

JOAO PRAZERES[1], (Student Member, IEEE), RAFAEL RODRIGUES[1], (Student Member, IEEE), MANUELA PEREIRA[2], AND ANTONIO M. G. PINHEIRO[1], (Senior Member, IEEE)

[1]Instituto de Telecomunicações, Universidade da Beira Interior, 6201-001 Covilhã, Portugal
[2]NOVA LINCS, Universidade da Beira Interior, 6201-001 Covilhã, Portugal

Corresponding author: Antonio M. G. Pinheiro (pinheiro@ubi.pt)

This work was supported by FCT/MECI through National Funds and EU Funds of the Instituto de Telecomunicações under Grant UID/50008.

ABSTRACT The quality evaluation of three deep learning-based coding solutions for point cloud geometry, notably ADLPCC, PCC GEO CNNv2, and PCGCv2, is presented. The MPEG G-PCC was used as an anchor. Furthermore, LUT SR, which uses multi-resolution Look-Up tables, was also considered. A set of six point clouds representing landscapes and objects was used. As point cloud texture has a great influence on the perceived quality, two different subjective studies that differ in the texture addition model are reported and statistically compared. In the first experiment, the dataset was first encoded with the identified codecs. Then, the texture of the original point cloud was mapped to the decoded point cloud using the *Meshlab* software, resulting in a point cloud with both geometry and texture information. Finally, the resulting point cloud was encoded with G-PCC using the `lossless-geometry-lossy-atts` mode, while in the second experiment the texture was mapped directly onto the distorted geometry. Moreover, both subjective evaluations were used to benchmark a set of objective point cloud quality metrics. The two experiments were shown to be statistically different, and the tested metrics revealed quite different behaviors for the two sets of data. The results reveal that the preferred method of evaluation is the encoding of texture information with G-PCC after mapping the texture of the original point cloud to the distorted point cloud. The results suggest that current objective metrics are not suitable to evaluate distortions created by machine learning-based codecs. Finally, this paper presents a study on the compression performance stability of the tested machine learning-based codecs using different training sessions. The obtained results show that the tested codecs revealed a high level of stability across all training sessions for most of the content, although some undesirable exceptions may be found.

INDEX TERMS Point clouds, machine learning, quality evaluation, coding.

I. INTRODUCTION

Point cloud technology is nowadays an established technique for three-dimensional data representation. A wide range of applications may employ point cloud technology, including virtual, augmented, and mixed reality; 3D printing; automation and robotics; computer graphics and gaming; and medical applications, among others.

In point clouds, objects or scenes are represented by a set of Cartesian coordinates (x, y, z), with each containing a list of attributes, such as RGB color components, reflectance values, normal vectors, or physical sensor information. They provide an accurate representation of objects or scenes from any viewing position or distance, thus making them a very powerful representation model. However, an accurate representation of a building or an artifact may require several million points and their associated attributes, leading to huge amounts of information. Hence, there is a need for efficient

The associate editor coordinating the review of this manuscript and approving it for publication was Ayman El-Baz.

point cloud coding solutions. The main contributions of this paper are as follows:

- A comparison between two different methods of adding texture to point clouds encoded by learning based solutions that only encode geometry in order to benchmark them. Two subjective evaluations were conducted with each method, and the result of both was compared.
- An analysis on the stability of machine learning-based coding solutions across different training sessions. This methodology follows the model of a previous study [1], aiming to study the reproducibility of learning based point cloud coding solutions.

For the comparison between evaluations, the MPEG G-PCC [2] was used as a benchmarking anchor for all studies. LUT SR, a fractional super-resolution method for G-PCC reconstructed point clouds [3], was also included. It should be emphasized that the main goal is to assert the best way to evaluate point cloud coding solutions that encode only geometry. The selected learning based solutions for this study were three DL-based codecs that only compress the geometry information. For a subjective study, texture information is required since it is highly important for the subjective quality definition. However, the method used for adding texture might be controversial, as it may introduce texture artifacts not created by the codecs under evaluation [4], [5]. For that reason, two different models were tested.

In the first model, the texture is compressed with G-PCC, and a balance between the compression rate of the texture and geometry is established. In the second model, the texture information was mapped from the original point clouds without further encoding onto the coded geometry. These two models are compared in order to understand which was the best model to add the texture information vital to an appropriate subjective study. While most previous studies only considered point cloud representations of small objects, point cloud representations of landscapes were included as well.

It is important to emphasize that the quality analysis could be done with point clouds without any texture [6]. However, during JPEG Pleno development [7], it was observed that the texture has a strong influence on subjective evaluations, and point cloud visualization without textures leads to results that are difficult to analyze. Moreover, the emerging deep learning codecs produce distortions that are extremely different from the ones created by the typical codecs (projection-based and octree-based). The typical subjective evaluation models were developed before the emergence of such technology, and their performance should be analyzed to verify which model is more adequate to evaluate this new generation of codecs.

Finally, this paper also provides a study on the stability of machine learning-based coding solutions across different training sessions, following the method of a previous study [1]. The remainder of this paper provides a short description of the state of the art in point cloud technology, notably in subjective quality evaluation, objective quality computation, and coding solutions, including the most popular DL-based solutions. The two subjective quality evaluation experiments are then described and analyzed, followed by a study on the performance of the selected objective metrics. In Section IV, the stability of the three DL-based solutions is analyzed. A conclusion section finalizes the paper.

II. STATE OF THE ART

A. POINT CLOUD CODING

The most traditional coding model for point clouds is based on the octree pruning method [8]. Recently, MPEG defined the Geometry-based Point Cloud Compression (G-PCC) [9] based on the octree point cloud representation. G-PCC also defines the trisoup method based on surface reconstruction for geometry compression. The point cloud attributes are compressed either with RAHT, the lifting transform, or the predictive transform. For this study, only the octree method is considered for geometry encoding and the lifting transform for texture encoding. It has been previously shown that subjects tend to prefer the lifting transform [10] over the RAHT algorithm. This method was selected in another evaluation [11], presenting good results.

Another trend for point cloud coding is the encoding of point cloud projections, which can be coded by any image coding codec. MPEG also explored that approach, resulting in the Video-based Point Cloud Compression (V-PCC) [2] defined for dynamic point clouds. V-PCC relies on HEVC (and more recently on VVC) to encode 2D projections of a given point cloud. Despite being developed for dynamic point clouds, its intra-coding has been revealed to be the most efficient for static point cloud coding [11], [12].

Following the good performance in image coding, several machine learning-based coding solutions for point clouds have been recently proposed [13], [14], [15], [16], [17], [18], [19], [20], [21]. These solutions usually cause distortions that are quite different from those caused by common codecs, which typically create holes in point cloud surfaces. Hence, there is a need to analyze the reliability of the quality models for learning-based codecs.

The DL-based codecs selected for this study are in concordance with an MPEG document [22]: Multiscale Point Cloud Geometry Compression (PCGC) [16], Deep Point Cloud Geometry Compression [18], and Adaptive Deep Learning Point Cloud Coding (ADLPCC) [19]. The codecs were compared against the MPEG anchor G-PCC [9] (V.14), using the octree mode, and also an evolution of G-PCC, LUT SR [3].

PCGCv2 [16] performs block-wise multi-resolution encoding. The point cloud is downsampled three times, and the encoding is done recurring to the Inception Residual Network [23]. At the bottleneck, the geometry coordinates are encoded with G-PCC, and entropy coding is used for the attributes. The decoding branch architecture mirrors the encoder. The implementation available at[1] was used.

[1] available at https://github.com/NJUVISION/PCGCv2

Deep Point Cloud Geometry Compression [18] learns an encoding function from three sequential convolution layers. The first two use ReLU activation. The latent representation of the third label is quantized through element-wise integer rounding and then compressed through a combination of algorithms. The decoding architecture mirrors the encoding. The output of the last layer is converted to the distorted point cloud using element-wise minimum, maximum, and rounding functions. The implementation available at[2] was used.

ADLPCC [19] partitions the point cloud into regular-sized 3D blocks. Several models separately code those blocks. The codec contains an autoencoder (AE) and a variational autoencoder (VAE) with three convolutional layers of both encoding and reconstruction, with sigmoid and ReLU activations, respectively. The implementation available at[3] was used.

Moreover, the LUT SR Look-Up Tables [3] solution is also considered. Based on G-PCC, it creates a hierarchical tree-like dictionary, mapping the occupancy relationships between downsampled geometry and the reference. A second downsampling is performed, storing the occupancy in the dictionary as well as the neighborhood configuration. The point cloud is then upsampled by applying nearest-neighbor interpolation to find all the possible child nodes of the input point cloud. The resulting geometry is obtained by following the respective dictionary entries. The implementation available at[4] was used.

B. POINT CLOUD SUBJECTIVE QUALITY EVALUATION

Several point cloud quality evaluation studies have been proposed, considering different coding methodologies and setups. Several studies established quality models for geometry-only encoding methods, such as octree-based [4], [6], [24], [25], graph-based [24], and projection-based encoding [4]. Su et al. [26] carried out a subjective evaluation of MPEG test models V-PCC and also S-PCC and L-PCC, which were earlier proposals that led to the final version of G-PCC. In the same year, Alexiou et al. also reported on an early subjective evaluation of G-PCC and V-PCC [10], before these were standardized. The previously mentioned study [11] reports a subjective quality evaluation of MPEG Point Cloud codecs using a 2D visualization setup. An initial quality study of DL-based point cloud coding quality was presented, targeting machine learning codecs [27]. Liu et al. [28], created a database for point cloud quality assessment, evaluating distortions from V-PCC, G-PCC, and other distortion types.

Subjective evaluation using augmented or virtual reality (AR/VR) environments has been previously researched [29], [30], [31], [32]. Alexiou et al. proposed PointXR [33], a toolbox for visualization and subjective evaluation of point clouds in VR environments. Recently, a subjective quality evaluation conducted in a 3D environment was presented in [34]. The obtained results were compared to a previous study using 2D displays [12] and showed no statistical differences. Crowdsourcing methodologies have also been studied [5] as a method of subjective evaluation. The participants were given the option of downloading the subjective evaluation content or accessing an online server and conducting the evaluation on a web browser. The two types of subjective evaluation revealed a very high level of statistical similarity. A subjective evaluation was conducted using a light field display and compared to an evaluation conducted with a 2D display [35]. The evaluations were highly correlated but presented statistical differences, and the authors concluded that no benefits were gained from using a light field display.

C. POINT CLOUD OBJECTIVE QUALITY EVALUATION

Objective quality metrics aim at accurately predicting the visual quality of content representations and may be used to set up codecs for an improved quality of experience without the need for subjective studies. These may be classified as image-based or model-based metrics specifically developed for point cloud quality evaluation [36]. Image-based metrics, such as PSNR and SSIM, operate directly on representative 2D views.

The aforementioned study [11] tested a group of point-based metrics and concluded that point-to-point and point-to-plane metrics [37] using the Mean Squared Error (MSE) were the best performing ones and provided a good representation of the subjective evaluation. Later, a benchmarking study using a 2D experimental setup for the subjective assessment, which included a broader selection of objective quality metrics, was reported [38], with PCQM and PointSSIM showing the best performances in terms of correlation with the Mean Opinion Score (MOS). Moreover, the image metric Multiscale Structural Similarity Index (MS-SSIM) computed over the video generated for the 2-D visualization of the point cloud also revealed a good representation of the subjective evaluation. Recently, no reference metrics are also being proposed [28], [39], [40], [41]. These metrics use features extracted from the distorted point cloud and deep learning technology to evaluate the compression quality.

For this paper, a set of objective metrics was considered, namely the MSE PSNR D1 and MSE PSNR D2 [37], the Point Cloud Structural Similarity (PointSSIM) Metric [42], the Point Cloud Quality Metric (PCQM) [43], the Point to Distribution Metric [44], the Reduced Reference Point Cloud Metric [45] and the GraphSIM [46] metric. These metrics are widely used in subjective evaluation, and they usually provide a good representation of subjective results [38]. The MSE PSNR D1 and D2 allow evaluations of the performance of the coding solutions regarding geometry, and the remaining metrics evaluate the performance based on both geometry and texture information. In the following, a short description of these metrics is presented.

[2]available at https://github.com/mauriceqch/pcc_geo_cnn_v2
[3]https://github.com/aguarda/ADLPCC
[4]https://github.com/digitalivp/PCC_LUT SR

1) POINT-TO-POINT - MSE PSNR D1 [37]

The MSE PSNR D1 metric measures geometric distortions by computing the Euclidean distance between every point b_k in the distorted input and the nearest corresponding point a_i in the reference point cloud, $E(a_i, b_k) = |(\overrightarrow{v}_{b_k}^{a_i})|_2$, and the MSE considering $E(a_i, b_k)$ is taken. The final output is given by the PSNR as follows: $D1 = 10 \cdot \log_{10}\left(\frac{max^2}{MSE}\right)$. The available MPEG implementation was used[1].

2) POINT-TO-PLANE - MSE PSNR D2 [37]

The D2 metric considers the projection of the error vector $\overrightarrow{v}_{b_k}^{a_i}$ along the surface normal of the nearest neighbor a_i (N_{a_i}). The MSE considering the projected errors $E(a_i, b_k) = \left|\overrightarrow{v}_{b_k}^{a_i} \cdot N_{a_i}\right|$ is taken, followed by PSNR. The available MPEG implementation was used[1].

3) POINT CLOUD STRUCTURAL SIMILARITY (POINTSSIM) [42]

This metric measures the statistical dispersion of attributes (either present or estimated) such as geometry information, color, normal vectors, or curvature. For each point p, a similarity index $S_Y(p)$ between the reference (X) and the distorted point cloud (Y), considering the K nearest neighbors, is given by:

$$S_Y(p) = \frac{|F_X(q) - F_Y(p|}{\max\{|F_X(q), F_Y(p)|\}} \tag{1}$$

where F_X and F_Y are the feature values of the reference and distorted point clouds, respectively. The final quality score S_Y is obtained by pooling all points.

$$S_Y = \frac{1}{N_p}\sum_{p=1}^{N_p} S_y(p)^K. \tag{2}$$

The implementation available online[5] was used, with the covariance (COV) as an estimator and $K = 12$ [42]. Also, color attributes were used, as they led to the best results in [42].

4) POINT CLOUD QUALITY METRIC [43]

The Point Cloud Quality Metric (PCQM) considers both geometry features based on the local mean curvature and color features computed on the LAB2000HL perceptual color space. Before feature computation, all points of the reference point cloud $p^X \in X$ are projected onto the 3D quadric surface subtended by the distorted point cloud Y, computed on a neighborhood of the closest point in Y. Each individual feature f_i takes into account the analogous values of p and its corresponding projected point $\hat{p}$. The final quality index is given by a weighted linear combination, $\text{PCQM} = \sum_{i\in S} w_i f_i$, where S is the set of indices of features and w_i are their associated weights. These are obtained by optimizing the linear model via logistic regression. The implementation made available by the authors[6] was used with the recommended weights.

5) POINT-TO-DISTRIBUTION [44]

This metric provides a joint geometry and color quality score ($P2D$) by computing the *Mahalanobis* distance between a point in point cloud X and the distribution of points in its K nearest neighborhood in point cloud Y. For geometry ($P2D - G$) and YUV color components ($P2D - C_m$), the distances are computed in two directions, i.e., reference to distorted and vice versa, with both results averaged across all points. The maximum between the two computations is then considered. The joint distance ($P2D - JGC_m$) is also obtained by averaging $P2D - G$ and $P2D - C_m$, and a final $P2D$ quality score is given by $P2D = \log_{10}\left(1 + \frac{1}{P2D - JGC_m}\right)$. In this work, the implementation was made available by the authors[7] was used, which considers only the luminance (Y) channel.

6) REDUCED-REFERENCE POINT CLOUD METRIC (PCM_{RR}) [45]

PCM_{RR} is based on features derived from geometry, luminance, and normal attributes. Notably, the mean, standard deviation, median, mode, entropy, energy, and sparsity are computed from the attributes and the occurrence histogram of both geometry and luminance. For normal attributes, the angular similarity θ between the normal of a given point and its K-nearest neighborhood is considered, along with the probability histogram of θ averaged across $k \in K$. The normal-based feature vector includes the mean of means, mean of standard deviations, mean of medians, standard deviation of means, entropy, energy, and sparsity. The final quality score is given by $\text{PCM}_{\text{RR}}. = \sum_i w_i d_i$, where d_i is the distance between the reference and distorted features with weights w_i, which are obtained by training on a point cloud set, maximizing the Pearson Linear Correlation Coefficient. The metric made available online[8] was used.

7) GRAPHSIM [46]

This metric first establishes a 3D keypoint skeleton $\vec{P}_S$, by resampling the reference point cloud through high-pass graph filtering. For each keypoint, a graph is constructed for both the reference and the distorted point clouds, connecting it to neighboring points below a certain threshold of the Euclidean distance. The gradient mass (m_g), the gradient mean (μ_g) and the gradient variance (σ_g^2) and the co-variance (c_g) are computed, with Euclidean distance-based weights being used to compute m_g and σ_g^2. Similarity measures are obtained for the color attributes and pooled channel-wise. This is done by aggregating it across all color components using the

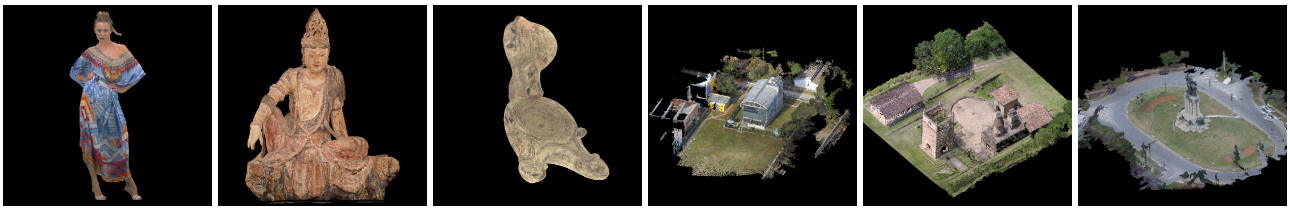
[1] http://mpegx.int-evry.fr/software/MPEG/PCC/mpeg-pcc-dmetric/tree/master

[5] https://github.com/mmspg/pointssim

[6] https://github.com/MEPP-team/PCQM

[7] https://github.com/AlirezaJav/Point_to_distribution_metric

[8] https://github.com/cwi-dis/PCM_RR



(a) Longdress (b) Guanyin (c) Romanoillamp (d) Citiusp (e) IpanemaCut (f) Ramos

FIGURE 1. Point Cloud test set.

TABLE 1. Point cloud characteristics.

Point Cloud	Sparsity (K=20)	Colour Gamut Volume	Y Deviation	Cb Deviation	Cr Deviation
Longdress	1.730	1.5%	0.114	0.060	0.065
Guanyin	1.748	1.3%	0.144	0.025	0.027
Romanoillamp	2.204	1.5%	0.094	0.020	0.012
Citiusp	1.788	4.03%	0.150	0.041	0.022
IpanemaCut	1.685	2.45%	0.161	0.032	0.022
Ramos	1.588	2.35%	0.125	0.029	0.014

following equation:

$$S_{\vec{s}_k} = \frac{1}{\gamma} \sum_{C} \gamma C \cdot |S_{\vec{s}_k, C}| \tag{3}$$

where γC is the pooling factor of the color channel. The metric is computed by averaging the different keypoint similarity contributions $S_{\vec{s}_k}$. The implementation available online[9] was used.

III. QUALITY ASSESSMENT OF DEEP LEARNING-BASED CODECS

In this section, details concerning both quality assessment models are provided. Section III-A explains the experimental setup and procedure for both evaluations and the conclusions drawn from the analysis of the results. Section III-B discusses the performance of the selected objective metrics in predicting the scores of the subjective evaluation.

A. SUBJECTIVE QUALITY EVALUATION

1) EXPERIMENTAL SETUP

Two subjective quality evaluations were carried out at the test laboratory of the Image and Video Technology Group of the Universidade da Beira Interior. Both subjective quality evaluations used the same set of six point clouds (Fig. 1), which includes three objects: frame 1300 of the *Longdress* dynamic point cloud available at the JPEG Pleno database,[10] *Guanyin* from the EPFL Dataset, and *Romanoillamp*, and three landscapes: *Citiusp*, *IpanemaCut*, and *Ramos* from the Univ. Sao Paulo Database.[11] Table 1 shows the selected point cloud sparsity, color gamut volume, and standard deviations of the YCbCr color channels. The sparsity is defined as the average distance between each point and its 20 nearest neighbors, averaged over the entire point cloud. The color gamut volume is defined as the volume of the convex hull of the distribution of color points in the YCbCr color space. The characteristics of Table 1 reflect a suitable degree of diversity for evaluating the coding solutions.

For the first subjective quality evaluation (*Evaluation 1*), the point cloud data was encoded using the codecs described in Section II-A, targeting five different encoding rates, ranging from poor quality (R01) to high quality (R05). As these codecs only encode geometry, there is a need to add texture to the distorted geometry, as it plays a very important role in quality perception. Hence, the texture information from the reference point cloud was mapped onto the distorted geometry. The resulting point clouds were then encoded with G-PCC using the `lossless-geometry-lossy-atts` mode. To achieve this, the QP parameter (which controls texture encoding) was set to $QP = \{0.25, 0.5, 0.75, 0.875, 0.9375\}$, for R01 to R05 (using the lifting transform), and the positionQuantizationScale (pQs), which controls geometry encoding, is set to 1. This ensures that no further artifacts are introduced by G-PCC in the distorted geometry. A total of 17 subjects (12 males and 5 females, ages 18–58 (24.7±8.3)) participated in *Evaluation 1*.

In the second subjective quality evaluation (*Evaluation 2*), the data preparation was similar, but instead of encoding texture using G-PCC, the texture of the reference point clouds was mapped directly onto the decoded geometry. A total of 17 subjects (11 males and 6 females, ages 21-32 (24.8±2.8)) participated in *Evaluation 2*.

[9] https://github.com/NJUVISION/GraphSIM
[10] http://plenodb.jpeg.org/pc/8ilabs
[11] http://uspaulopc.di.ubi.pt

TABLE 2. Point size of each point cloud for visualization in the subjective test.

	ADLPCC					PCC_GEO_CNN					PCGC				
Content	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05
Longdress	7	4	3			8	6	5	3		12	10	9	3	
Guanyin	7	4	3			8	6	5		3	12	10	9	3	
Romanoillamp	15	6		5		10	8		7	6	15	14	13	8	7
Citiusp	3					3					3				
IpanemaCut	3					3					3				
Ramos	3					3					3				
	G-PCC					**LUT SR**									
	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05					
Longdress	8	4	3			3									
Guanyin	12	10	9	3		3									
Romanoillamp	10	6	5			5									
Citiusp	12	6	5	3		3									
IpanemaCut	12	6	5	3		3									
Ramos	12	6	4	3		3									

For both experiments, the texture was mapped using *Meshlab*.[12] The software maps the color using the nearest neighbor algorithm. For the point cloud without texture information, the nearest neighbor of the point cloud with texture information is identified. The color of the nearest neighbor is then attributed to that point.

The point size was adjusted heuristically, as shown in Table 2. This is important to create continuous surfaces, thus avoiding perceptual effects caused by transparency [4], [6]. However, it should also be carefully adjusted so that the point size does not mask the distortion artifacts created by the codecs. For content representing landscapes, only G-PCC required an increase in the point size for lower bitrates. All other codecs maintained good surface integrity for the visualization using the standard point size without a relevant influence on the perceived quality. In the case of content representing objects, most codecs require adjustment, especially at lower bitrates. This was not verified for LUT SR.

For both evaluations, videos depicting the reference and distorted point clouds were prepared. A point cloud view was captured for each 1° degree rotation using PCL Visualizer [8], completing a full rotation around the vertical axis. For point clouds depicting objects, the frontal view was chosen as the initial view. Furthermore, in the case of landscapes, the viewing point was rotated to an angle of 43° degrees with the xy plane, allowing a top view visualization, as their frontal view is not suitable for subjective evaluation.

The full sequences of frames were then rendered with FFMPEG,[13] using the H.264 codec [47] at 30 fps, resulting in 12-second videos. To ensure that no compression was applied, the CRF (Constant Rate Factor) and q were set to 0. The `libx264rgb` option was used to prevent any RGB to YUV conversion. The subjective test setup used a 31.1-inch Eizo ColorEdge CG318-4K with a full resolution of 4096 × 2160 and followed the specifications in [48].

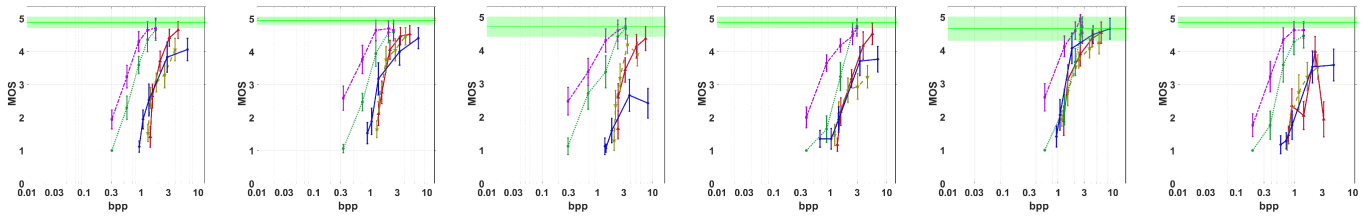
Before starting any quality evaluation, all participants were shown eight training videos with encoded versions of two point clouds not included in the test set, i.e., *Airplane* from the PointNet Database and *Villalobospark*, from the University of Sao Paulo Database. This allowed an adaptation to typical encoding distortions, to the evaluation scale, and also to the user interface. During the evaluation, each participant was shown a unique, randomized sequence of videos of the distorted point clouds and the respective reference, side by side. Reference/reference pairs were also included for hidden reference evaluation. Distortions of the same point cloud were never shown one after the other. Moreover, half of the subjects performed the subjective quality evaluation with the reference on the right, whereas the other half had the reference on the left.

A Double Stimulus Impairment Scale method was adopted, with the subjects being prompted to evaluate the quality of the distorted point cloud in comparison to the provided reference according to a five-level rating scale (1: very annoying, 2: annoying, 3: slightly annoying, 4: perceptible but not annoying, 5: imperceptible). After the subjective test, the MOS for all stimuli was computed.

2) RESULTS AND DISCUSSION

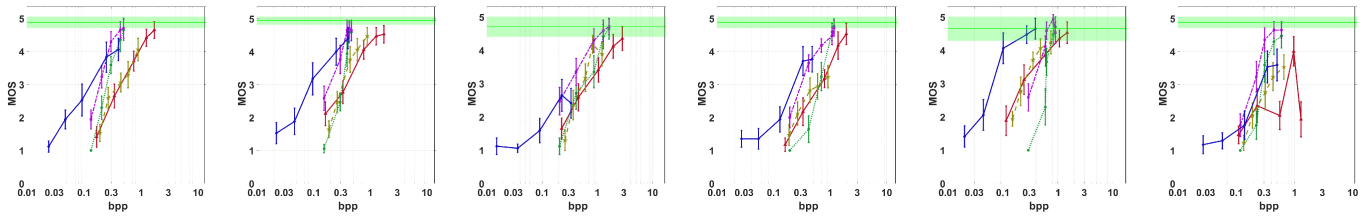
In *Evaluation 1*, since the texture of the original point cloud was mapped to the distorted one and then encoded using G-PCC, the texture bitrate is added to the geometry bitrate. Fig. 3 also refers to the results of *Evaluation 1*, but only the geometry bitrate is considered. In *Evaluation 2*, represented in Fig. 4, the texture information was directly mapped onto the distorted geometry and no texture coding was applied, hence the bitrate is relative to geometry only. Although the scores do not follow a Gaussian distribution, their 95% Confidence Interval (CI) was computed, assuming a Student's t-distribution. The horizontal green line at the top of each plot refers to the MOS for the hidden references, whereas the green bar around it represents its 95% CI. The vertical black line on the right side of each plot represents the lossless encoding with G-PCC. This was computed to assure that the tested bitrates were not larger than the lossless bitrate of G-PCC. It should be noted here that G-PCC and LUT

[12] https://www.meshlab.net
[13] https://ffmpeg.org/



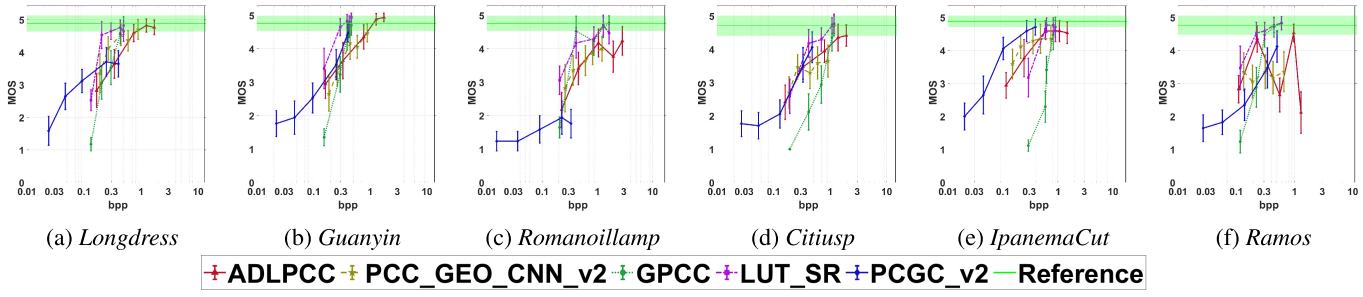
(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

FIGURE 2. MOS vs bpp with 95% CI for ***Evaluation 1*** (texture encoded with G-PCC). The bitrate results from geometry and texture.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

FIGURE 3. MOS vs bpp with 95% CI for ***Evaluation 1*** (texture encoded with G-PCC). Here, bpp refers to the geometry bitrate only.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

FIGURE 4. MOS vs bpp with 95% CI for ***Evaluation 2*** (texture directly mapped onto the distorted geometry). Here, bpp refers to the geometry bitrate only.

SR were tested for similar bitrates, allowing a more direct comparison. For the deep learning-based codecs, i.e., PCC GEO CNNv2, PCGCv2, and ADLPCC, the resulting bitrates are highly dependent on their training, so it is not possible to select the bitrates freely. Nevertheless, these are directly comparable within their selected range of bitrates and are simultaneously comparable to the higher bitrates of G-PCC and LUT SR.

It can be concluded that none of the machine-learning codecs can reach the performance of G-PCC when the encoded texture is added. The bitrate achieved by them is always superior to G-PCC, except for the *IpaemaCut* point cloud. PCGCv2 only reaches a MOS similar to the reference for the *IpanemaCut* (R04 and R05) and the *Guanyin* (R05) point clouds. However, it seems to have slightly better performance at low bitrates than ADLPCC and PCC GEO CNNv2 for some of the tested content, notably *Longdress*, *Guanyin*, and *IpanemaCut*. The *Romanoillamp* point cloud is an outlier to this general behavior, as PCGCv2 performed quite poorly with it. The MOS did not reach 3 at any bitrate, which may be related to a lack of suitable data in the training set. In the case of ADLPCC, the *Ramos* point cloud exhibits strange behavior, as the higher bitrate results in very low performance. This might also be caused by a lack of suitable training data or overfitting. Another strange behavior is observed for the higher bitrate of the *Romanoillamp* point cloud encoded with PCC GEO CNNv2. The plot shows a slight decrease in MOS in that case. Since it is a very small decrease, it can be concluded that it is due to the randomness of the test sequence, as it can also influence the scores given by the subjects.

Fig. 3 shows the influence of the geometry bit rate alone, without the influence of the texture that is coded with G-PCC in the case of the deep-learning codecs, shown in Fig. 2. It allows observing how effective the deep learning solution was in compressing the subjective quality. PCGCv2 exhibits the best performing DL-based solution overall, except for *Romanoillamp*, where PCC GEO CNNv2 achieved the best performance.

Fig. 4 shows that most codecs achieve very competitive results, outperforming G-PCC for almost every point cloud. The only exception is observed for the *Romanoillamp*

(a) R01 (b) R03 (c) R05

FIGURE 5. *Romanoillamp* encoded with PCGCv2.

(a) R01 (b) R03 (c) R05

FIGURE 6. *Ramos* encoded with ADLPCC (crop).

(a) R01 (b) R03 (c) R05

FIGURE 7. *IpanemaCut* encoded with PCC GEO CNNv2 (crop).

point cloud. The best performing codec is PCGCv2. As expected, the same drop in the MOS at the highest rate is observed for the *Ramos* point cloud encoded with ADLPCC, as well as the poor overall evaluation of the *Romanoillamp* point cloud encoded with PCGCv2. It should also be noted that for the *Guanyin* point cloud, PCGCv2 outperformed LUT SR, with similar scores at lower bitrates. From the plots, it can also be concluded that *Evaluation 1* usually presents lower confidence intervals than *Evaluation 2*, revealing a more stable grading from the subjects.

Figs. 5 to 7 show examples of three point clouds, each encoded with a different codec and at three quality levels, from lower (left) to higher (right).

Fig. 5 shows several artifacts present across all bitrates for *Romanoillamp* encoded with PCGCv2. These artifacts are most likely derived from the downsampling and upsampling operations. When downsampling a point cloud at the encoder stage, some information regarding that process can be lost, negatively impacting the upsampling process on the decoder side. Fig. 6 shows that a large part of the column is missing in the R05 rate of the*Ramos* point cloud encoded

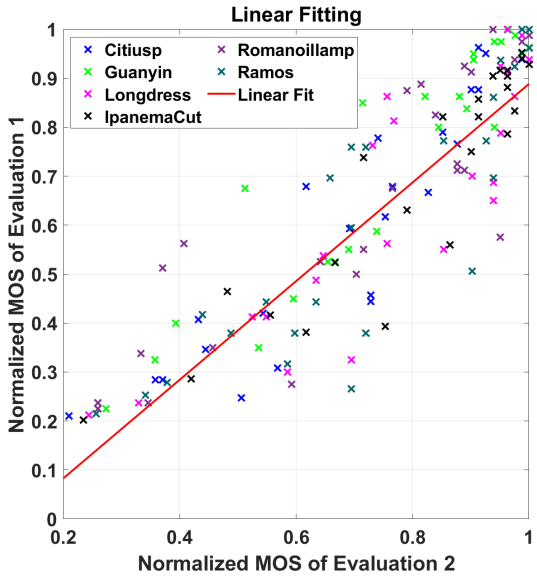


FIGURE 8. MOS scatter plot (geometry-only vs. geometry and texture) with the respective linear fitting. The MOS scores were normalized between 0 and 1.

TABLE 3. Statistical analysis between both conducted subjective quality evaluations.

Linear Fitting			
PCC	SROCC	RMSE	OR
0.894	0.908	0.740	0.139

with ADLPCC. The codec relies on dividing the point cloud into blocks, followed by coding them separately. Some blocks are flagged as empty blocks after the lossy process, even if they exist in the original point cloud. Because of that, some parts of the point clouds might be missing. This is common in the lower bitrates but should not happen for the higher quality levels.

Fig. 7 shows a cropped area with several artifacts in the tower present across bitrates for the *IpanemaCut* encoded with PCC GEO CNNv2. PCC GEO CNNv2 uses convolutional layers to encode and decode the point clouds. The resulting artifacts are most likely due to the bit-wise and element-wise rounding that occurs in the final layer of the encoding and decoding architecture. Sometimes the artifacts appear in the higher bitrates, which is caused by the lack of appropriate training data to handle such point clouds.

A very important part of this work is to analyze how effective the two subjective models for quality evaluation are. For that, the statistical similarity of the two subjective evaluations is analyzed. Fig. 8 shows the linear fitting between the MOS scores of both evaluations and Table 3 shows the Pearson Correlation Coefficient (PCC), Spearman Rank Order Correlation Coefficient (SROCC), the Root Mean Square Error (RMSE), and the Outlier Ratio (OR) values between them using linear fitting, as recommended in ITU-T P.1401 [49]. The results reveal some statistical similarity,

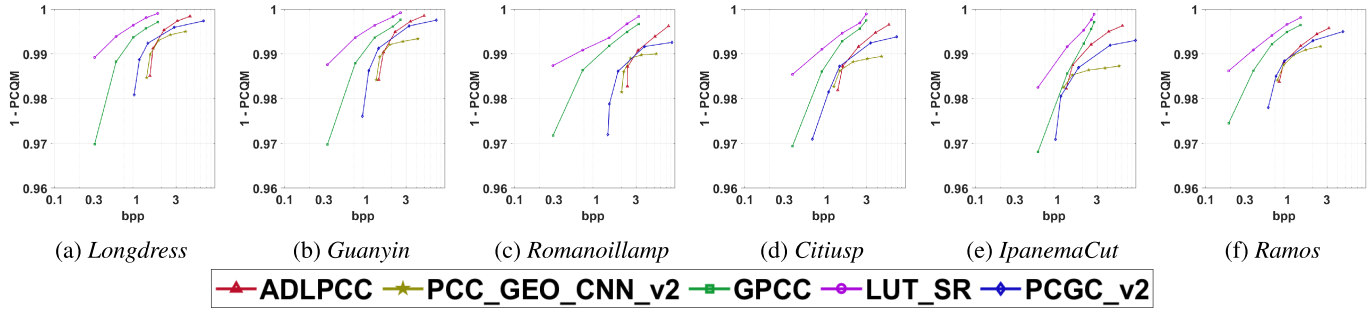


FIGURE 9. 1 - PCQM vs bpp (geometry + texture).

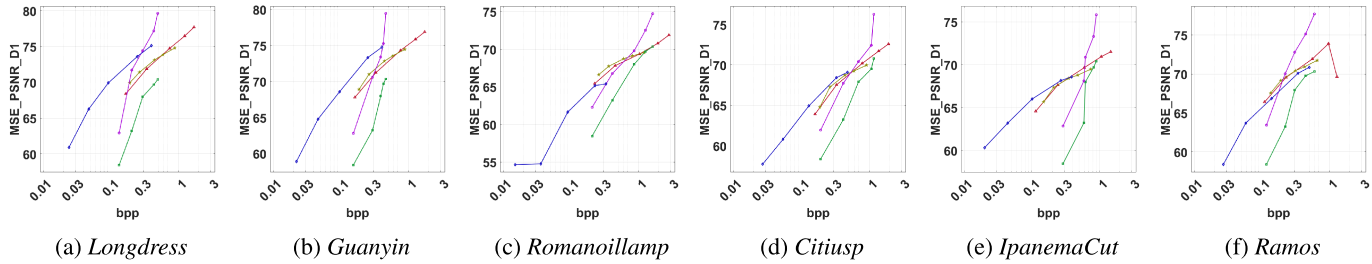


FIGURE 10. MSE PSNR D1 vs bpp (geometry only).

but the two subjective evaluations do not seem to provide the same results.

To further assess the statistical similarity between the two evaluations, a Kruskal-Wallis one-way analysis followed by a multiple comparison test [50] was performed. It was concluded that there are statistical differences between the two subjective evaluations (p-value = 7.479E-04 $<$ 0.05).

At this stage, it is important to try to understand which of the two subjective evaluation models is better. It is important to consider that mapping the texture onto the distorted geometry without any encoding (*Evaluation 2*) results in a subsampling process that is likely to cause aliasing. This may eventually influence the perceived quality. From the observation of the samples, it is observed that encoding the texture provides a better balance between geometry and texture quality, resulting in a more reliable subjective evaluation.

B. PERFORMANCE OF OBJECTIVE QUALITY EVALUATION

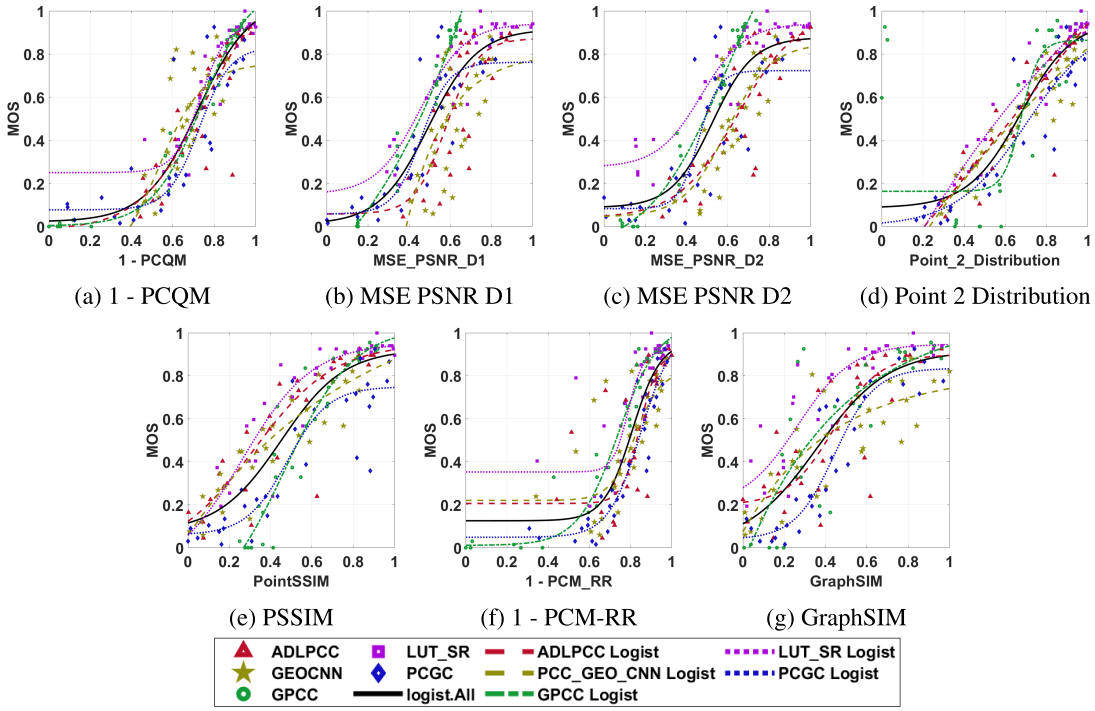
Subjective quality assessment provides the ground truth for the validation of objective quality metrics, considering the distortions produced by the tested codecs. Figs. 9 and 10 show the PCQM and MSE PSNR D1 metrics, respectively, plotted against the coding bitrates. In Fig. 9, the bitrates consider both geometry and texture, as PCQM is a joint perceptual metric. On the other hand, Fig. 10 considers only the geometry bitrate, as MSE PSNR D1 only uses the geometry information.

The plots in Fig. 9 indicate that PCQM tends to establish the same performance relations between different codecs as the subjective evaluations, although some exceptions can be identified easily. The PGCGv2 codec obtained results similar to the other codecs for the *Romanoillamp* point cloud, which is not identified in the subjective evaluation. This metric also failed to predict the unusual behavior observed with the *Ramos* point cloud encoded with ADLPCC at R05.

MSE PSNR D1 is represented in Fig. 10. It is observed that it does not follow the performance found in the subjective evaluations represented in Figs. 3 or 4. The metric shows that G-PCC is the worst performing codec, which is not the case in any evaluation. The metric predicts that PCGCvs is the best performing codec, which is contrary to what is shown in 2. The point clouds *IpanemaCut* and *Ramos* have some similarity between the metric MSE PSNR D1 and subjective results. The quality decrease at rate R05 for *Ramos* is predicted, but that does not happen for R03. The metric also did not predict the quality decrease for the rate R05 for *Romanoillamp* encoded with PCC GEO CNNv2.

To evaluate the performance of each metric, the usual benchmarking procedure [49], [51]. MOS predictions for each metric were obtained by logistic regression of the objective scores. Then, PCC, SROCC, RMSE, and OR were computed to measure the correlation between these and the subjective MOS results, as specified in [49]. For the metrics that depend on surface normals, the Cloud Compare Quadric Fitting with a radius of 5 [52] was used, as this value usually provides the best metric performances [53].

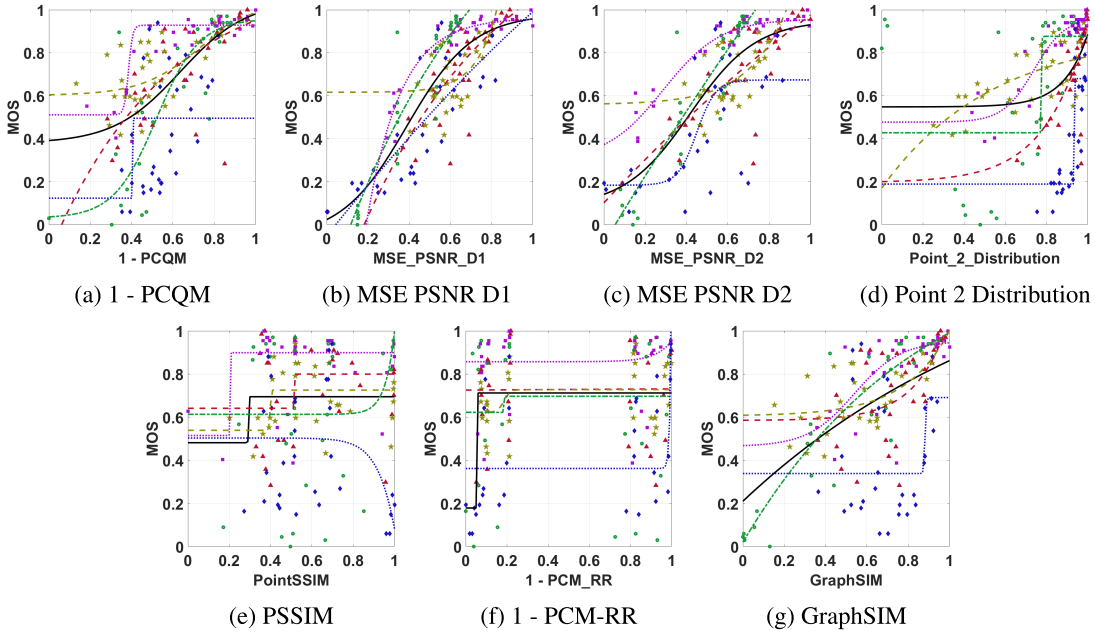
Figs. 11 and 12 show the normalized objective metric vs. normalized MOS plots *Evaluation 1* and *2*, respectively, and Table 4 summarizes the corresponding correlation measures



(a) 1 - PCQM (b) MSE PSNR D1 (c) MSE PSNR D2 (d) Point 2 Distribution

(e) PSSIM (f) 1 - PCM-RR (g) GraphSIM

FIGURE 11. Objective metric vs. Evaluation 1 MOS plots, with logistic regression curves (global and for each codec).



(a) 1 - PCQM (b) MSE PSNR D1 (c) MSE PSNR D2 (d) Point 2 Distribution

(e) PSSIM (f) 1 - PCM-RR (g) GraphSIM

FIGURE 12. Objective metric vs. Evaluation 2 MOS plots, with logistic regression curves (global and for each codec). The symbols are the same as Fig. 11.

for both experiments. In the table, the best values are shown in bold, and the second-best values are shown in italic.

PCQM shows the best correlation with the subjective MOS for *Evaluation 1* (PCC/SROCC of 0.899/0.903), while MSE PSNR D1 shows the best correlation for *Evaluation 2* (PCC/SROCC of 0.834/0.774). The geometry-only metrics MSE PSNR D1 and MSE PSNR D2 are the only objective metrics with similar performance in the two evaluations, although both PCC and SROCC indicate a poor representation of the subjective quality. Joint objective metrics fail to predict the compression quality when the reference texture information is mapped onto the distorted geometry (*Evaluation 2*), supporting the preference for the encoded texture model used in *Evaluation 1*.

TABLE 4. Correlation of the objective metrics with the subjective quality evaluation results. The best values are shown in bold, and the second best values are shown in italic.

		Evaluation 1			
Metric	Type	PCC	SROCC	RMSE	OR
MSE PSNR D1	*FR, GEO*	0.806	0.782	0.184	0.753
MSE PSNR D2	*FR, GEO*	0.821	0.796	0.177	0.813
PointSSIM	*FR, COL*	*0.859*	*0.857*	*0.159*	*0.720*
Point 2 Distribution	*FR, GEO + COL*	0.851	0.828	0.164	0.640
PCM-RR	*FR, GEO + COL*	0.834	0.834	0.172	0.727
GraphSIM	*FR, GEO + COL*	0.800	0.799	0.186	0.780
PCQM	*FR, GEO + COL*	**0.899**	**0.903**	**0.137**	**0.573**
		Evaluation 2			
MSE PSNR D1	*FR, GEO*	**0.834**	**0.774**	**0.152**	**0.720**
MSE PSNR D2	*FR, GEO*	*0.777*	*0.740*	*0.174*	*0.793*
PointSSIM	*FR, COL*	0.188	0.143	0.271	0.920
Point 2 Distribution	*FR, GEO + COL*	0.437	0.472	0.249	0.873
PCM-RR	*FR, GEO + COL*	0.408	0.323	0.252	0.900
GraphSIM	*FR, GEO + COL*	0.560	0.573	0.229	0.907
PCQM	*FR, GEO + COL*	0.634	0.700	0.214	0.787

Quite unexpected was the very low performance of PCQM. Even though it was the best performing metric for *Evaluation 1*, it did not achieve the same performance as a previous study [12]. The reference implementation assigns fixed weights to each feature based on a linear-optimization algorithm [43], and a point cloud database was used to compute them. The reasons for this behavior might be the fact that those values are not suitable for learning-based codecs.

IV. PERFORMANCE STABILITY OF DEEP LEARNING-BASED CODECS

In the following section, the performance stability of the tested DL-based codecs is studied for three different training sessions with similar conditions. The evolution of the codecs throughout the learning process is first analyzed using the *Guanyin*, *Romanoillamp*, and *Citiusp* point clouds and by computing the MSE PSNR D1 at each training epoch. Finally, the coding performance of the resulting operating points, as well as the publicly available implementations, is assessed using PCQM and MSE PSNR D1, considering the six point clouds referred to in Section III-A.

The three DL-based codecs studied here are trained three times, keeping all the training conditions. The global loss function depends on the distortion of the encoded point clouds and the encoding bitrate. The encoding bitrate is estimated differently for each codec. PCGCv2 estimates the distortion from the Binary Cross-entropy loss function (BCE), $BCE = -\frac{1}{N}\sum_i (x_i \log(p_i) + (1 - x_i \log(1 - p_i)))$, where x_i is the true binary occupancy value of voxel i, and p_i is its occupancy probability output given by the model. PCC GEO CNNv2 and ADLPCC use a focal variation of the BCE to address imbalances between empty and occupied voxels in more sparse point clouds, defined as:

$$\begin{cases} -\alpha(1-x)^{\gamma}\log p, & x = 1 \\ -(1-\alpha)x^{\gamma}\log(1-p), & x = 0 \end{cases} \tag{4}$$

A. PCGCV2 MODEL TRAINING

The PCGCv2 codec implementation and the training datasets are available online.[14] The model was trained with densely sampled data from the ShapeNet database [54]. The final training set was obtained by random rotation and quantization with 7-bit precision and a randomized number of points.

In the original paper, different coding bitrates are targeted by varying the rate-distortion tradeoff parameter λ between 0.75 and 16. The code made available defines the global loss function as, $J = \alpha D + \beta R$ where D is the distortion and R is the coding bitrate. In this experiment, three different λ values were tested, namely $\lambda = \{16, 4, 0.75\}$. The β parameter was kept with the value 1 so that we have $J = \lambda D + R$.

For faster convergence, the learned weights with $\lambda = 16$ were used to initialize the training for both $\lambda = 4$ and $\lambda = 0.75$, as recommended in [16]. For each λ, the model was trained for 50 epochs with a constant learning rate of 10^{-5}. This training process was repeated three times under similar conditions.

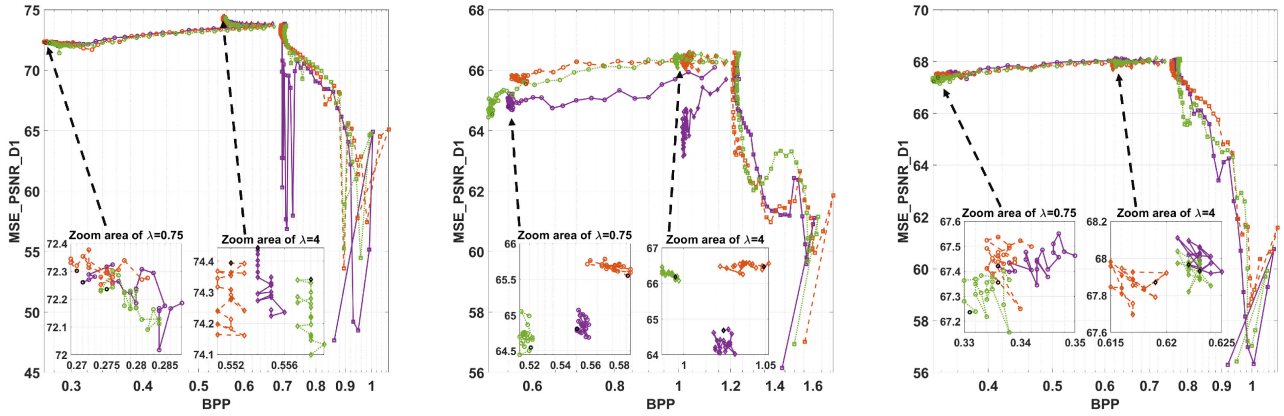
The evolution of MSE PSNR D1 vs. the coding bitrate throughout the model training is shown in Fig. 13. The zoomed areas show the final epochs of each Rate-Distortion tradeoff. PCGCv2 reveals a good level of stability for most point clouds, notably *Citiusp*. The MSE PSNR D1 metric shows similar behavior across the three training sessions. However, even in these cases, some instability may be observed, for example, in intermediate epochs with $\lambda = 16$, particularly at higher encoding bitrates for *Guanyin Citiusp*. The encoding process of *Romanoillamp* point cloud shows an extremely high level of instability. It can also be observed that in this particular case, the training results do not converge to a stable operating point, except for $\lambda = 16$. The codec reveals a good level of stability regarding the encoding performance of the other tested point clouds. The bitrates converge to similar operating points across all training sessions. In conclusion, the codec usually reveals a good level of stability, but there is a possibility that other training data will produce situations like *Romanoillamp*, where the performance of the codec will depend on the training.

B. PCC GEO CNNV2 MODEL TRAINING

The authors of PCC GEO CNNv2 train four individual models for each Rate-Distortion tradeoff given by $J = \lambda D + R$ [18]. They chose four values for λ, notably 3×10^{-4}, 10^{-4}, 5×10^{-5}, 2×10^{-5}. In the provided implementation,[15] an additional value is considered, $\lambda = 10^{-5}$, which was also included here. This experiment followed the sequential training approach in [18], with successively decreasing values of λ. The trained weights for λ_{i-1} were used to initialize the training for λ_i).

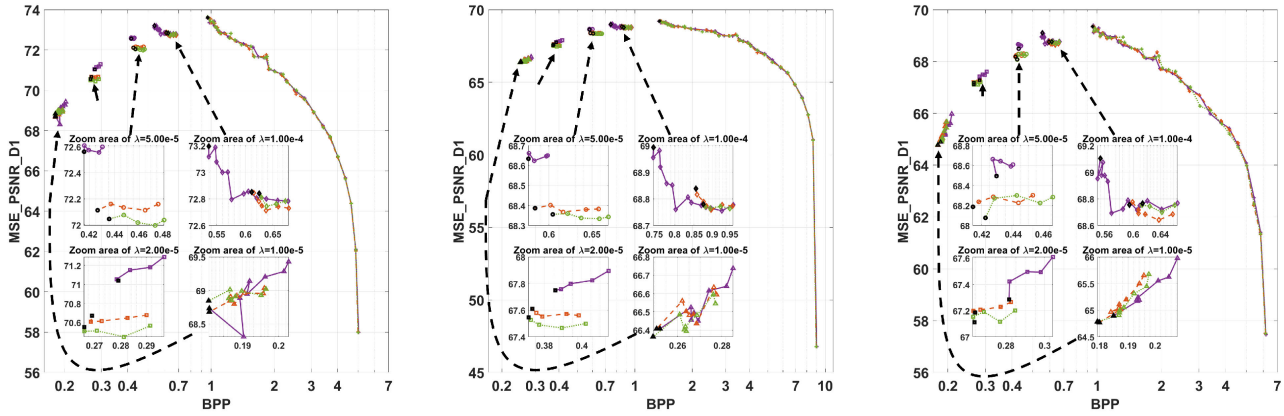
The models were trained on a subset of the ModelNet40 [55] dataset. First, the mesh data is voxelized with a resolution of $512 \times 512 \times 512$, and the 200 largest point clouds

[14] available at https://github.com/NJUVISION/PCGCv2
[15] available at https://github.com/mauriceqch/pcc_geo_cnn_v2



(a) *Guanyin* (b) *Romanoillamp* (c) *Citiusp*

FIGURE 13. MSE PSNR D1 vs. bpp plots for PCGC, trained with $\alpha = \{16, 4, 0.75\}$.



(a) *Guanyin* (b) *Romanoillamp* (c) *Citiusp*

FIGURE 14. MSE PSNR D1 vs. bpp plots for PCC GEO CNNv2, trained with $\lambda = \{3 \times 10^{-4}, 10^{-4}, 5 \times 10^{-5}, 2 \times 10^{-5}, 10^{-5}\}$.

are selected. Then, the point clouds are divided into blocks with a resolution of $64 \times 64 \times 64$, and the 4000 largest blocks are selected. For each value of λ, the model was trained for 500 steps, with early stopping if the loss did not improve for more than 4 validation steps.

Fig. 14 shows the evolution of MSE PSNR D1 for three training sessions of PCC GEO CNNv2. The zoomed areas show the final epochs of each Rate-Distortion tradeoff. This codec shows a very high level of stability. However, for intermediate rates, i.e., with $\lambda = 1 \times 10^{-4}$, 5×10^{-5}, and 2×10^{-5}, the MSE PSNR D1 values of the final resulting models are slightly different across training sessions. In practice, no case in which the final coding result is highly dependent on the training session was identified.

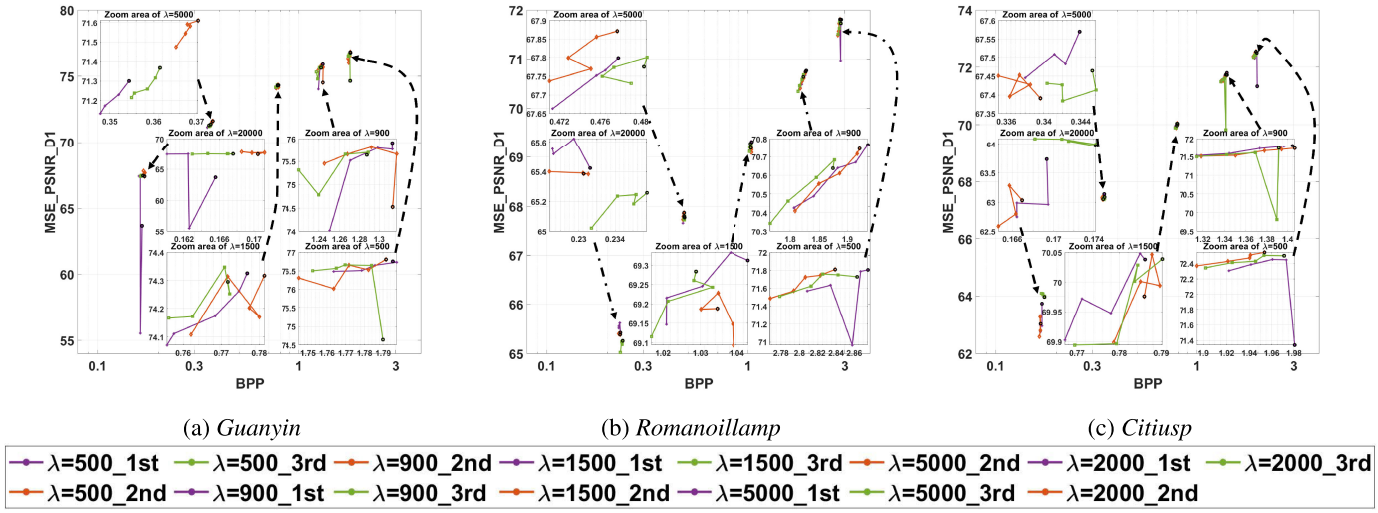
C. ADLPCC MODEL TRAINING

The global loss function of ADLPCC[16] is given by $J = D + \lambda R$, where the coding rate R is estimated during training as the summed entropy of its autoencoder and variational autoencoder latent representations.

In order to obtain several Rate-Distortion tradeoff points, different λ values are considered, thus varying the weight of the rate. The model was trained with a dataset consisting of JPEG and MPEG point clouds [19], with $\lambda = \{500, 900, 1500, 5000, 20000\}$. For each value of λ, the codec was trained with $\alpha = \{0.5, 0.6, 0.7, 0.8, 0.9\}$, which is a parameter of the BCE focal loss function (Eq. 4), which allows choosing the best performing model considering the characteristics of the point cloud, such as its sparsity.

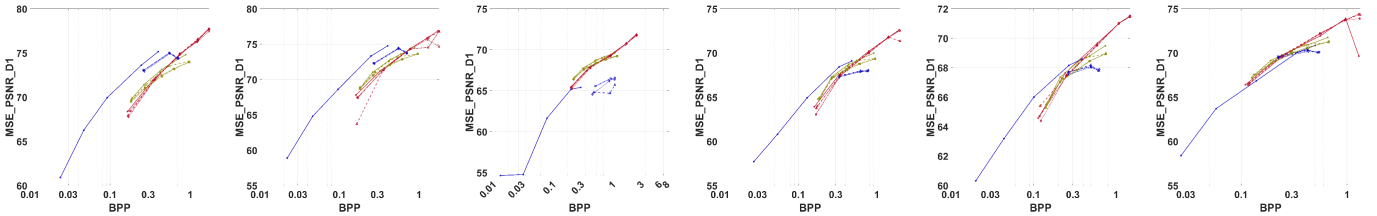
Fig. 15 shows the plots for each epoch of the ADLPCC codec across training. The zoomed areas show the final epochs of each Rate-Distortion tradeoff. The codec shows a high degree of stability, as most encoding steps show little variation across the epochs. One notable exception should be noted, namely for the *Guanyin* point cloud when trained with $\lambda = 20000$. The first train revealed a sudden drop in one of the intermediate epochs. The rest of the training process for that specific λ showed little variation in quality across bitrates.

[16] https://github.com/aguarda/ADLPCC



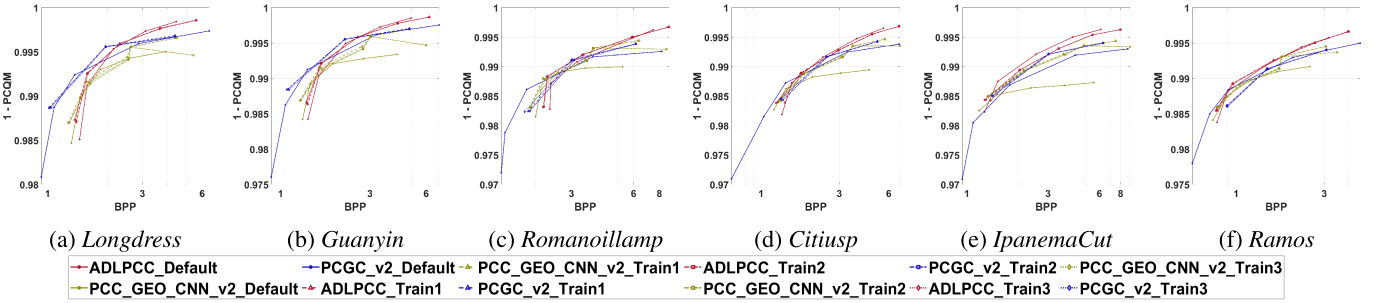
(a) *Guanyin* (b) *Romanoillamp* (c) *Citiusp*

FIGURE 15. MSE PSNR D1 vs. bpp plots for ADLPCC, trained with $\lambda = \{500, 900, 1500, 5000, 20000\}$.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

FIGURE 16. MSE PSNR D1 plots for each of the defined operating points for each codec.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

FIGURE 17. 1 - PCQM plots for each of the defined operating points for each codec.

D. FINAL COMMENTS ON THE CODECS PERFORMANCE STABILITY

Figs. 16 and 17 show the performance of the three codecs that resulted from the three training sessions. It is important to understand that the three codecs have different performances for different training sessions. The figures also depict the performance of the default implementations of ADLPCC, PCGCv2 and PCC GEO CNNv2.

The ADLPCC is the one that depends less on the training session but has a performance decrease in the higher bit rate for three point clouds and the second higher bit rate for one point cloud. It is important to emphasize that the higher bit rates are the ones that provide an acceptable quality, and they might be the ones most commonly used in practical applications, which makes this inconsistency a problem. It should be noted that the three different training sessions did not produce the quality drop at the higher rate.

PCC GEO CNNv2 has one training situation that consistently leads to much better performance in the middle bit rates than the other training situations. This level of variation establishes some unreliability in the codecs performance.

The PGCCv2 is the most stable codec and presents the best performance on the middle bit rates for the *Longdress* and *Guanyin* point clouds. The three training sessions also produced poor results for the *Romanoillamp* point cloud. The larger bit rate can result in lower performance and is not reliable. This is caused by the training model that starts to obtain this bit rate and then adjusts the cost function to the remaining operating points. It has the problem of not reaching near-perceptual lossless qualities. It should be noted that the authors do not specify the lambda trade-off that they use in the current implementation. As such, some results will vary depending on the default codec and the training sessions conducted.

V. CONCLUSION

A study on the performance of machine learning-based codecs for static point clouds was reported, notably, PCGCv2, PCC GEO CNNv2, and ADLPCC.

Because these codecs are geometry-only codecs, two different models that included texture were tested. Based on previous studies and observations, texture is essential for allowing reliable subjective evaluation. From the results, it was concluded that the two tests are statistically different. Furthermore, it was concluded that encoding the texture provides a more reliable subjective evaluation than just mapping the original one to the resulting geometry. Several objective metrics were computed and correlated with the results of both evaluations. The PCQM reveals the best representation of the subjective results for *Evaluation 1*, and MSE PSNR D1 showed the best results for *Evaluation 2*. The point cloud objective metrics did not provide a good representation for the subjective evaluation, where texture was just mapped on the decoded point clouds without any compression.

Finally, the stability of different training sessions was analyzed for the three codecs. Although in most cases the performance slightly changes for different training sessions, there were cases where a significant quality variation resulted. This is highly undesirable, as it results in a reduction in the point cloud coding performance, depending on the training session. Moreover, the performance can depend too much on the content, which is also undesirable (like in the case of *Romanoillamp* for PCGCv2).

It is important to emphasize the difficulty of DL-based codecs in reaching near-perceptual lossless coding, which potentially makes their use difficult to be adopted by the community. In some cases, an undesired decrease in quality is observed with the increase in bit rate, mainly for the high bit rates. This might be caused by a lack of training data or overfitting. DL-based codecs define different decoders for each bit rate. One possible solution is to define multiple encoders and decoders for high bit rates and use the one that provides the best bit-rate distortion ratio. Eventually, the decoder and encoder settings could be defined in the metadata, and non-default decoders could even be included in the metadata. Furthermore, these non-default codecs could also be compressed. MPEG recently created a standard for neural network compression that could be used. However, the bit-rate distortion consequences of using it in this scenario still need to be researched and evaluated.

REFERENCES

[1] J. Prazeres, R. Rodrigues, M. Pereira, and A. M. G. Pinheiro, "On the stability of point cloud machine learning based coding," in *Proc. 10th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Sep. 2022, pp. 1–6.

[2] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Trans. Signal Inf. Process.*, vol. 9, no. 1, p. 13, 2020.

[3] T. M. Borges, D. C. Garcia, and R. L. de Queiroz, "Fractional super-resolution of voxelized point clouds," *IEEE Trans. Image Process.*, vol. 31, pp. 1380–1390, 2022.

[4] L. A. Da Silva Cruz, E. Dumic, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation: Towards a definition for test conditions," in *Proc. 11th Int. Conf. Quality Multimedia Exp. (QoMEX)*, Jun. 2019, pp. 1–6.

[5] S. Perry, L. A. Da Silva Cruz, E. Dumic, N. H. Thi Nguyen, A. Pinheiro, and E. Alexiou, "Comparison of remote subjective assessment strategies in the context of the JPEG pleno point cloud activity," in *Proc. IEEE 23rd Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2021, pp. 1–6.

[6] 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 *Proc. 10th Int. Conf. Quality Multimedia Exp. (QoMEX)*, May 2018, pp. 1–6.

[7] P. Astola, L. A. Da Silva Cruz, E. A. Da Silva, T. Ebrahimi, P. G. Freitas, A. Gilles, K.-J. Oh, C. Pagliari, F. Pereira, C. Perra, S. Perry, A. M. G. Pinheiro, P. Schelkens, I. Seidel, and I. Tabus, "Jpeg pleno: Standardizing a coding framework and tools for plenoptic imaging modalities," *ITU J., ICT Discoveries*, vol. 3, no. 1, pp. 1–15, Jun. 2020.

[8] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 1–4.

[9] K. Mammou, P. A. Chou, D. Flynn, and M. Krivokuća, *G-PCC Codec Description V2*, Standard JTC1/SC29/WG11 N18189, ISO/IEC, Jan. 2019.

[10] E. Alexiou, I. Viola, T. M. Borges, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A comprehensive study of the rate-distortion performance in MPEG point cloud compression," *APSIPA Trans. Signal Inf. Process.*, vol. 8, no. 1, p. 27, 2019.

[11] 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 *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 3428–3432.

[12] J. Prazeres, M. Pereira, and A. Pinheiro, "Quality analysis of point cloud coding solutions," in *Proc. Electron. Imag. Symp.*, vol. 34, Jan. 2022, pp. 225–226.

[13] A. F. R. Guarda, N. M. M. Rodrigues, and F. Pereira, "Point cloud coding: Adopting a deep learning-based approach," in *Proc. Picture Coding Symp. (PCS)*, Nov. 2019, pp. 1–5.

[14] A. F. R. Guarda, N. M. M. Rodrigues, and F. Pereira, "Deep learning-based point cloud geometry coding with resolution scalability," in *Proc. IEEE 22nd Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2020, pp. 1–6.

[15] J. Wang, H. Zhu, Z. Ma, T. Chen, H. Liu, and Q. Shen, "Learned point cloud geometry compression," 2019, *arXiv:1909.12037*.

[16] J. Wang, D. Ding, Z. Li, and Z. Ma, "Multiscale point cloud geometry compression," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2021, pp. 73–82.

[17] M. Quach, G. Valenzise, and F. Dufaux, "Learning convolutional transforms for lossy point cloud geometry compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 4320–4324.

[18] M. Quach, G. Valenzise, and F. Dufaux, "Improved deep point cloud geometry compression," in *Proc. IEEE 22nd Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2020, pp. 1–6.
[19] A. F. R. Guarda, N. M. M. Rodrigues, and F. Pereira, "Adaptive deep learning-based point cloud geometry coding," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 2, pp. 415–430, Feb. 2021.
[20] J. Pang, M. A. Lodhi, and D. Tian, "GRASP-net: Geometric residual analysis and synthesis for point cloud compression," in *Proc. 1st Int. Workshop Adv. Point Cloud Compress., Process. Anal.*, Oct. 2022, pp. 11–19.
[21] J. Wang, D. Ding, Z. Li, X. Feng, C. Cao, and Z. Ma, "Sparse tensor-based multiscale representation for point cloud geometry compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 9055–9071, Jul. 2022.
[22] *Performance Analysis of Currently AI-Based Available Solutions for PCC*, Standard JTC1/SC29 WG7, ISO/IEC, Oct. 2022.
[23] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2017, vol. 31, no. 1, pp. 4278–4284.
[24] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of compressed point clouds," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6.
[25] E. Alexious, A. M. Pinheiro, C. Duarte, D. Matković, E. Dumić, L. A. Da Silva Cruz, L. G. Dmitrović, M. V. Bernardo, M. Pereira, and T. Ebrahimi, "Point cloud subjective evaluation methodology based on reconstructed surfaces," *Proc. SPIE*, vol. 10752, pp. 160–173, Sep. 2018.
[26] H. Su, Z. Duanmu, W. Liu, Q. Liu, and Z. Wang, "Perceptual quality assessment of 3D point clouds," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 3182–3186.
[27] J. Prazeres, R. Rodrigues, M. Pereira, and A. M. G. Pinheiro, "Quality evaluation of machine learning-based point cloud coding solutions," in *Proc. 1st Int. Workshop Adv. Point Cloud Compress., Process. Anal.*, Oct. 2022, pp. 57–65.
[28] Q. Liu, H. Su, Z. Duanmu, W. Liu, and Z. Wang, "Perceptual quality assessment of colored 3D point clouds," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 8, pp. 3642–3655, Aug. 2023.
[29] E. Alexiou, E. Upenik, and T. Ebrahimi, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6.
[30] 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 *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2020, pp. 127–136.
[31] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 4, pp. 828–842, Apr. 2017.
[32] S. Subramanyam, I. Viola, J. Jansen, E. Alexiou, A. Hanjalic, and P. Cesar, "Subjective QoE evaluation of user-centered adaptive streaming of dynamic point clouds," in *Proc. 14th Int. Conf. Quality Multimedia Exp. (QoMEX)*, Sep. 2022, pp. 1–6.
[33] E. Alexiou, N. Yang, and T. Ebrahimi, "PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *Proc. 12th Int. Conf. Quality Multimedia Exp. (QoMEX)*, May 2020, pp. 1–6.
[34] J. Prazeres, M. Pereira, and A. M. G. Pinheiro, "Subjective quality evaluation of point clouds with 3D stereoscopic visualization," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 2861–2865.
[35] D. Lazzarotto, M. Testolina, and T. Ebrahimi, "On the impact of spatial rendering on point cloud subjective visual quality assessment," in *Proc. 14th Int. Conf. Quality Multimedia Exp. (QoMEX)*, Sep. 2022, pp. 1–6.
[36] G. Lavoué, M. C. Larabi, and L. Vávsa, "On the efficiency of image metrics for evaluating the visual quality of 3D models," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 8, pp. 1987–1999, Aug. 2016.
[37] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3460–3464.
[38] D. Lazzarotto, E. Alexiou, and T. Ebrahimi, "Benchmarking of objective quality metrics for point cloud compression," in *Proc. IEEE 23rd Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2021, pp. 1–6.
[39] Q. Yang, Y. Liu, S. Chen, Y. Xu, and J. Sun, "No-reference point cloud quality assessment via domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 21147–21156.
[40] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, and G. Zhai, "No-reference quality assessment for 3D colored point cloud and mesh models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7618–7631, Nov. 2022.
[41] Y. Liu, Q. Yang, Y. Xu, and L. Yang, "Point cloud quality assessment: Dataset construction and learning-based no-reference metric," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 2s, pp. 1–26, Jun. 2023.
[42] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2020, pp. 1–6.
[43] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "PCQM: A full-reference quality metric for colored 3D point clouds," in *Proc. 12th Int. Conf. Quality Multimedia Exp. (QoMEX)*, May 2020, pp. 1–6.
[44] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "A point-to-distribution joint geometry and color metric for point cloud quality assessment," in *Proc. IEEE 23rd Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2021, pp. 1–6.
[45] I. Viola and P. Cesar, "A reduced reference metric for visual quality evaluation of point cloud contents," *IEEE Signal Process. Lett.*, vol. 27, pp. 1660–1664, 2020.
[46] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, "Inferring point cloud quality via graph similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 3015–3029, Jun. 2022.
[47] T. Wiegand and G. J. Sullivan, "The H.264/AVC video coding standard [standards in a nutshell]," *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 148–153, Mar. 2007.
[48] B. Series, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation ITU-R BT*, vol. 500, no. 13, 2012.
[49] *Methods, Metrics and Procedures for Statistical Evaluation, Qualification and Comparison of Objective Quality Prediction Models*, Standard I.-T. P.1401, Int. Telecommun. Union, Recommendation, Jul. 2012.
[50] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, Dec. 1952.
[51] M. V. Bernardo, A. M. G. Pinheiro, P. T. Fiadeiro, and M. Pereira, "Image quality under chromatic impairments," *ACM Trans. Appl. Perception*, vol. 14, no. 1, pp. 1–20, Aug. 2016.
[52] P. Bo, R. Ling, and W. Wang, "A revisit to fitting parametric surfaces to point clouds," *Comput. Graph.*, vol. 36, no. 5, pp. 534–540, Aug. 2012.
[53] *JPEG Pleno PC Exploration Study 4 Results*, Standard WG1M89044, ISO/IEC JTC1/SC29/WG1, Sep. 2020.
[54] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "ShapeNet: An information-rich 3D model repository," 2015, *arXiv:1512.03012*.
[55] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1912–1920.

JOAO PRAZERES (Student Member, IEEE) received the Graduate and master's degrees in electrical and computer engineering from the Universidade da Beira Interior (UBI), Covilhã, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree. He has been deeply involved in the JPEG PLENO Point Cloud Coding activity. He received the Best Paper Award in 3-D Imaging and Applications of the Electronic Imaging Symposium 2022.

RAFAEL RODRIGUES (Student Member, IEEE) received the B.Sc. degree in biomedical sciences and the M.Sc. degree in electrical and computer engineering from the Universidade da Beira Interior (UBI), Portugal, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. He has also collaborated closely in many research efforts on image and video coding and quality assessment. His primary research interests include medical image processing and computer-aided diagnosis, with involvement on the EU COST Action BM1304–MYO-MRI. He is currently a Portuguese JPEG Member and a Qualinet (COST IC1003) and VQEG Member.

MANUELA PEREIRA received the B.S. degree in mathematics and computer science and the M.Sc. degree in computational mathematics from the University of Minho, Portugal, in 1994 and 1999, respectively, and the Ph.D. degree in signal and image processing from the University of Nice Sophia Antipolis, France, in 2004. She is an Associate Professor with the Computer Science Department, Universidade da Beira Interior, Portugal. Her main research interests include image and video coding, multimedia technologies standardization, signal processing for telecommunications, information theory, real-time video streaming, 3-D and 4-D imaging, and medical imaging.

ANTONIO M. G. PINHEIRO (Senior Member, IEEE) received the Licenciatura degree in electrical and computer engineering from IST, Lisbon, in 1988, and the Ph.D. degree in electronic systems engineering from the University of Essex, U.K., in 2002. He is an Associate Professor with the Universidade da Beira Interior (UBI); and a Researcher with the Instituto de Telecomunicações (IT), Portugal. He is a Portuguese Delegate of ISO/IEC JTC1/SC29 and the Communication Subgroup Chair of JPEG. He was the PC Co-Chair of QoMEX 2015, the Special Session Co-Chair of QoMEX 2016, and an Organizer of the tutorial in ACM Multimedia 2021 "Plenoptic Quality Assessment: The JPEG Pleno Experience." He is an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA.

5.3 Quality Analysis of Point Cloud Coding solutions

Quality analysis of point cloud coding solutions
J. Prazeres, Manuela Pereira, and Antonio Pinheiro
Electronic Imaging 34 (2022): 1-6. DOI:10.2352/EI.2022.34.17.3DIA-225

https://doi.org/10.2352/EI.2022.34.17.3DIA-225

Quality analysis of point cloud coding solutions

João Prazeres, Manuela Pereira, Antonio M. G. Pinheiro
Instituto de Telecomunicações and Universidade da Beira Interior, Covilhã, Portugal

Abstract

In this paper, a subjective quality based comparison between four point clouds codecs is presented. For that, a set of six point clouds was chosen. They were coded with four different point cloud encoding solutions, notably the MPEG V-PCC and G-PCC, a deep learning coding solution RS-DLPCC and also Draco, with different bit rates. A subjective test where the distorted and reference point clouds were rotated in a video sequence side by side followed by the quality evaluation, was conducted. Then the performance of a set of four point cloud objective quality metrics of he quality, was analysed using the subjective quality evaluation results. These metrics are usually reported as providing a good representation and are often used to evaluate compression solutions. In fact, the studied metrics tend to provide a good representation for V-PCC and G-PCC, an acceptable representation for RS-DLPCC, and a bad representation for Draco. It was also concluded that V-PCC is the best codec of the studied ones. The deep learning based solution still performs worst than the two MPEG codecs.

Introduction

In the modern world, 3D data capture and transmission became a common requirement for emerging technologies. Typical 3D information representation leads to huge amounts of data. Therefore, efficient methods of data compression are needed, in order to provide efficient transmission and storage of 3D data. Recently, point cloud technology has emerged as a very popular method for 3D data representation. A point cloud is a set of Cartesian coordinates(x, y, z), with a list of attributes associated to each element, such as a RGB component, reflectance information, physical sensor information or normal vectors. Point clouds contain a large amount of information, allowing an accurate representation of an object or scene. Hence, they are a very powerful visual representation model, extremely useful in VR/AR scenarios [22], computer graphics or 3D computer vision applications, between others.

If an accurate precision point cloud of a city or building, or even of an artefact is created, the resulting file can easily have several millions of points, with several features associated to each point. Since the representation of 3D data can contain a large amount of information, several solutions for point cloud compression have been researched. MPEG provided encoding solutions, notably V-PCC (Video Point Cloud Compression) and G-PCC (Geometry Point Cloud Compression) [7]. Deep Learning technology has been considered for image and video compression. Some works had also considered that possibility for point clouds compression [5, 6, 20, 23].

In this paper, the two MPEG codecs and a deep learning based solution proposed to the JPEG Pleno Point Cloud coding call for evidence, RS-DLPCC, are considered [20]. Furthermore, the DRACO codec [26] that has gain some popularity as a royalty free coding solution for Point Clouds and meshes was also considered.

This paper aims to compare the performance of the four codecs, using subjective and objective quality evaluation models. Several works have been published considering the quality evaluation of point clouds. In [9, 10], geometry only point clouds are considered. Compression artifacts using prior encoding schemes are evaluated in [11–13]. Current efforts account for a wider range of high-performing codecs, such as the ones reported in [7,14,15]. In [7] a quality model for point clouds is established. Apart a subjective evaluation using the MPEG codecs, the work also considers a set of point cloud metrics concluding that point to point and point to plane metrics [16] are the best performing ones and provide a good representation of the subjective evaluation. A subjective quality evaluation test where GPCC, V-PCC, RS-DLPCC and Draco codecs are compared was performed. We believe this is the first study that compares these four coding solutions. Finally, a comparison between the subjective results and objective models is also performed.

Brief Description of the tested codecs

Geometry Point Cloud Compression

G-PCC (Geometry Point Cloud Compression) [3] contains two methods of point cloud compression, an octree based method, and a trisoup based, method. For this study, only the octree method was considered. The octree compression method is regulated in the codec by the positionQuantizationScale (pQS) parameter. This parameter controls the number of divisions of the octree from the root, to each leaf node leading to a regular down sampling of the input clouds. Five different rates were coded for each content, ranging from low to near perfect quality levels, with bitrates ranging from 0.09 to 12 bits per point.

Table 1: G-PCC Parameters Example

Rate	QP	pQS
R01	46	0.125
R02	40	0.25
R03	34	0.5
R04	28	0.75
R05	22	0.875

Video Point Cloud Compression

The V-PCC (Video Point Cloud Compression) [4], presents a solution which projects the point cloud in a set of planes, and then encodes the projections in the 2D domain. Those projections contain texture, depth and an occupancy map, with the textures being encoded with legacy methods and the depth being encoded with 2D video encoding methods. The occupancy map represents the pixels containing meaningful information, and is encoded with

spatial quantization, combined with raster scanning and entropy encoding. The image projection sequence is encoded with the HEVC video codec. Five rates were chosen for each of the point clouds in the set, with bitrates ranging from 0.08 to 15.22 bits per point.

Table 2: V-PCC Parameters Example

Rate	Geometry QP	Texture QP	Occupancy Map
R01	36	47	4
R02	32	42	4
R03	28	37	4
R04	20	27	4
R05	16	22	2

Resolution Scalable Deep-Learning Point Cloud Compression (RS-DLPCC)

This codec uses a deep-learning approach to compress point clouds geometry [5], by using a latent representation of a point cloud, computed by an auto encoder framework. The scalability feature is made possible by interlaced block creation. The point cloud is divided into super-blocks, which are further divided by interlaced down sampling, resulting in up to eight interlaced blocks for each super-block, which are then coded separately, then enabling random-access. For each point cloud, four rates were chosen, with bitrates ranging from 0.34 to 25.88 bits per point.

This codec is likely to create some blocking artifacts due to the super blocks division.

After the geometry encoding, the color was transferred from the nearest neighbour of the original point cloud. The color for the recolored points is encoded with G-PCC, using the lossless geometry Octree coding mode, and the Predlift color encoder. The lossless Octree coding mode was chosen so that the (decoded) geometry is not changed, minimising the geometry coding effects on the color information from the G-PCC codec. This color information is then textured over the RS-DLPCC lossy decoded geometry.

Draco

Draco is a popular codec developed by Google. This codec uses KD-Tree [21] in order to efficiently organize the 3D data. Draco continuously splits the point cloud from the center, while also modifying the axes on each direction. Draco comes with four main parameters for controlling point cloud encoding. QP, which define the quantization bits for the position attributes, QT, which defines the quantization bits for the texture coordinate attribute, QN, which defines the quantization bits for the normal vector attribute, and QG, which defines the quantization bits for any generic attribute. Draco contains 32 levels of quantization (0 - 31) and 11 levels of compression. For this test, qp levels of 7, 9 and 10 were considered, which represent low, medium and high quality, respectively with the default compression level of 7. The coded point clouds resulted in bitrates ranging from 8.1 to 28.41 bpp.

Basically, Draco is a lossless codec. The parameter qp was used to control the bit rate, but basically it controls the precision of the representation. Reducing the precision (or somehow the resolution) of the point cloud representation allows the codec to be more efficient, resulting in lower bit rates. This was also the reason to consider 3 bit rates only. The resulting bit rates are much higher as qp does not reduce the number of points. It changes the point locations, reducing the representation precision, resulting at the same time lower bit rates.

Evaluation Methodology

Point Cloud Data Selection

For the comparative study, a set of six point clouds was used, containing geometry and texture information. The set consisted of a frame selected from the soldier and longdress dynamic point clouds available at [1], representing human figures. Frames 1300 and 0690 were selected for the longdress and soldier, respectively. Furthermore the point clouds rhetorician, guanyin, from EPFL dataset and point clouds romanoillamp, bumbameuboi, available at [2] were also selected. The later four point clouds represent cultural heritage artifacts. The selected point clouds are represented in figure 1. The full body point cloud redandblack (frame 1550) [1] was used for training prior to the subjective evaluation. The set was coded using V-PCC, G-PCC, RS-DLPCC and Draco with different bit rates.

Subjective evaluation

For all point clouds, a complete rotation over the vertical axis was applied. At each degree an image representing the point cloud view was extracted. These images were extracted using PCL Visualizer. The point cloud views were rendered as 12 second videos, and were displayed at 30fps and with 1920x1080 resolution. Videos were created using the FFMPEG software using a no video compression mode. To ensure no compression was applied to the extracted frames, the stream copy option in FFMPEG was used [8].

In some cases, the point size was changed to provide an improved visual representation. If holes appear in the point cloud the viewers will see the opposite part of the point cloud and that creates a very bad quality perception [10,12]. The change of point size is important to avoid this effect and to create continuous surfaces for the point cloud under observation. The point size values are represented in table 3 for each content and were obtained for the display used in this subjective test. The point clouds bumbameuboi and romanoillamp require a different solution from the remaining content. For point clouds coded with V-PCC, a modification was not required, so the default value of 1 was set. For all point clouds coded with G-PCC, the point size was set to 6 for R01 and 4 for R02. All the other rates were set to the default value. For the remaining cases the options are described in table 3 C

(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Figure 1: Point Cloud testing set.

Table 3: Point size for each content

V-PCC					
Content	R01	R02	R03	R04	R05
bumbameuboi	4	4	4	4	4
Guanyin	1	1	1	1	1
Longdress	1	1	1	1	1
Rhetorician	1	1	1	1	1
Romanoillamp	2	2	2	2	2
Soldier	1	1	1	1	1
G-PCC					
Content	R01	R02	R03	R04	R05
bumbameuboi	6	4	4	4	4
Guanyin	6	4	1	1	1
Longdress	6	4	1	1	1
Rhetorician	6	4	1	1	1
Romanoillamp	3	2	2	2	2
Soldier	6	4	1	1	1
RS-DLPCC					
Content	R01	R02	R03	R04	R05
bumbameuboi	-	20	9	8	7
Guanyin	-	6	4	1	1
Longdress	-	6	4	1	1
Rhetorician	-	6	4	1	1
Romanoillamp	-	7	3	2	2
Soldier	-	6	5	1	1
Draco					
Content	R01	R02	R03	R04	R05
bumbameuboi	6	-	4	-	4
Guanyin	6	-	2	-	1
Longdress	6	-	2	-	1
Rhetorician	6	-	2		1
Romanoillamp	6	-	2	-	2
Soldier	6	-	2	-	1

For the test, a Double Stimulus Impairment Scale was used. In this method, both the reference and the coded point cloud are shown to the subject. Then the subject is asked to evaluate each point cloud pair difference in a five-level rating scale (1 - very annoying, 2 - slightly annoying, 3 - annoying, 4 - perceptible, but not annoying, 5 - imperceptible). Prior to the evaluation, a sequence of four videos was shown to the subjects to help familiarizing with the evaluation. The redandblack point cloud was selected with four different levels of degradation. This point cloud was not included in the final test sequence. Additionally, hidden reference-reference pairs were included in the test sequence, to help verifying unusual behaviour in the evaluation. The same content was never shown twice in a row. To avoid biases, half the subjects were shown videos with the reference on the right and the codec content on the left, and vice-versa. All the tests were, conducted at subjective test laboratory of Image and Video Technology Group of Universidade da Beira Interior, using a 47 inch, FULL HD LG 47LA860V, with the test environment following the specifications in [24].

Six different point clouds were selected for the subjective quality evaluation of this test, based on the experience of JPEG and MPEG evaluation test sets. Both V-PCC and G-PCC codecs had five quality levels, while RS-DLPCC had four quality levels, and Draco had three quality levels. Taking the references in account, a total of 108 scores were obtained in each session.

Table 4: Subject Information

Males	Females	Overall	Age Span	Average age
10	6	16	21-33	24.75

Subjective Evaluation Scores

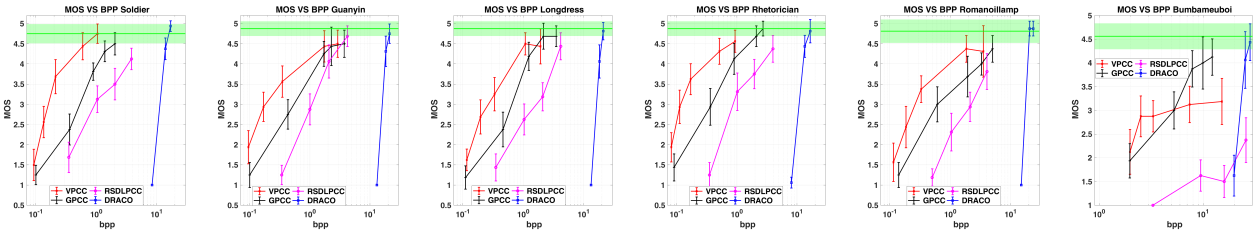
After the test, all the scores were aggregated, and the MOS for each content was computed. The bitrate, measured in bits per point (bpp), is calculated by taking the number of bits of a particular content, and dividing it by the number of points of the original content.

The MOS results are represented in figure 2. These figures also represent the Confidence Interval (considering a Gaussian distribution) The green line and the green horizontal bar represent the MOS and respective confidence interval obtained for the original point cloud (that was also in the test as hidden reference). This bar can be seen as a representation where transparent quality is reached. Although this is a simplistic approach, it is somehow representative that a given codec reaches unperceived distortions in case the MOS is inside this green horizontal bar. Moreover, can be observed: 1) The V-PCC in general provides the best quality scores, followed by G-PCC and the RS-DLPCC. Draco is the worst case, leading to much higher bit rates. The point cloud bumbameuboi is the exception to this regular behavior. This happens because this point cloud is rather sparse when compared with the others. This also reveals that further studies will be required in the future for sparse point clouds, that tend to be created with some acquisition technologies, like LIDARS.

Objective Evaluation

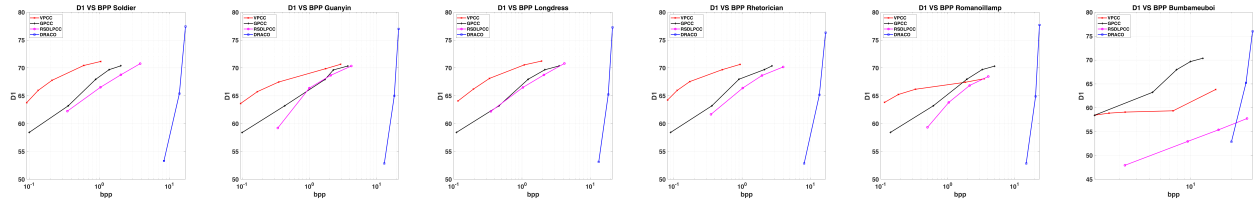
Four objective metrics were calculated:

- Point-to-point: This metric calculates the geometric distance of associated points between the reference and of the



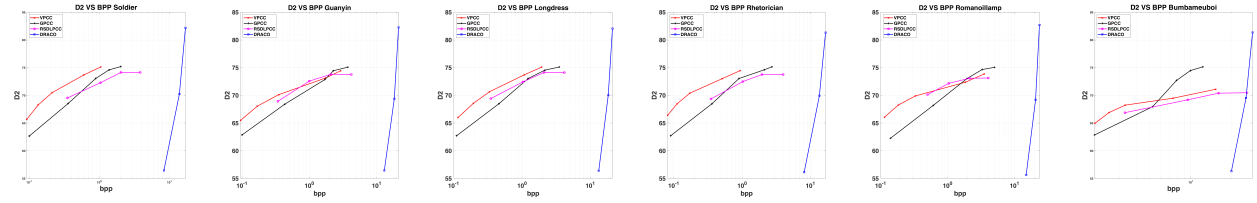
(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Figure 2: MOS with Confidence intervals (assuming a Gaussian distributiion) vs bpp



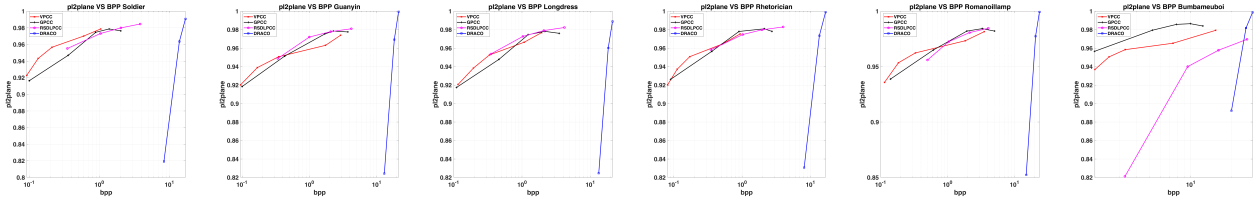
(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Figure 3: D1 vs bpp



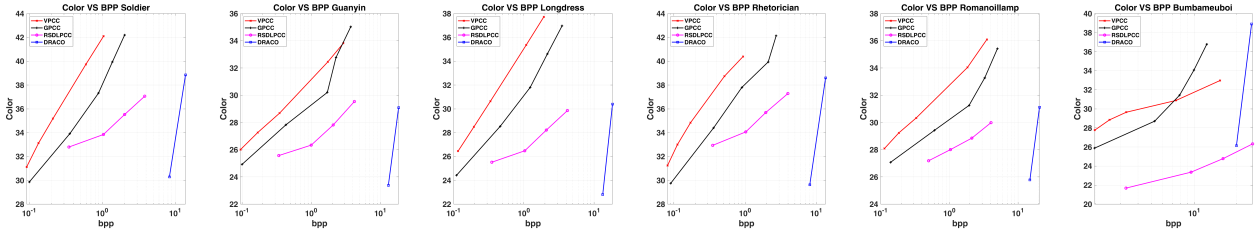
(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Figure 4: D2 vs bpp



(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Figure 5: pl2plane vs bpp



(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Figure 6: Color vs bpp

content under evaluation. Through nearest neighbour algorithm, the corresponding point belonging to the reference point cloud of each point of the distorted point cloud is found. Individual errors are computed based on Euclidean distance [16], followed by the aggregation mechanism.

- Point-to-plane: For every point of the content under evaluation, a point in the reference cloud is identified, through the nearest neighbour algorithm. A plane is fitted into the region centered on that point that is normal to the point under consideration. This plane is computed using quadric fitting in CloudCompare [18], with a radius of 20. The individual error of the point of the content under evaluation is the dimension of the normal vector to the plane and ends on that point. The final metric is the aggregation of these individual errors [16].
- plane-to-plane: For each point in the coded content, a point is identified using the nearest neighbour algorithm. Afterwards, considering the normal vectors to the planes for the reference and coded point cloud, the angular similarity is calculated. This is computed for each point [17]. This metric requires the planes normal to the vectors of both point clouds. The planes were computed using quadric fitting in cloud compare, with a radius of 20 [18].
- color: For every point in the codec point cloud, a point is identified, belonging to the reference cloud, through nearest neighbour algorithm. An individual error is computed based on Euclidean distance, and for color attributes, the MSE is calculated for the three color components, with a RGB to YCbCr conversion being made [25].

The plots of these metrics versus bit rate are represented respectively in figures 3, 4, 5 and 6. The highest bit rate of Draco resulted in infinite values for 6 and could not be represented.

From plots 3 to 5 we can observe that the lower performance found in the subjective evaluation (2) for the deep learning solution RS-DLPCC is not visually observed as it exhibits a very close performance to the G-PCC. While the metrics studied in this work provide a good representation of the subjective quality for the MPEG codecs, they did not reveal the same representation for the deep learning solution. Moreover, these metrics are not appropriate to evaluate the Draco encoder performance. Typically, deep learning solutions lead to different types of distortions which are not properly represented by the studied metrics.

Objective Metrics Benchmarking

In order to compare the objective measures with the subjective scores, the statistical measures proposed in [19] were calculated to measure the performance of each metric. Specifically, these are the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coeficient (SROCC), the Root-Mean Squared Error (RMSE) and the Outlier Ratio (OR). The prediction of the MOS for each objective metric was computed by applying a linear (no fitting) and a logistic fitting function on the objective scores.

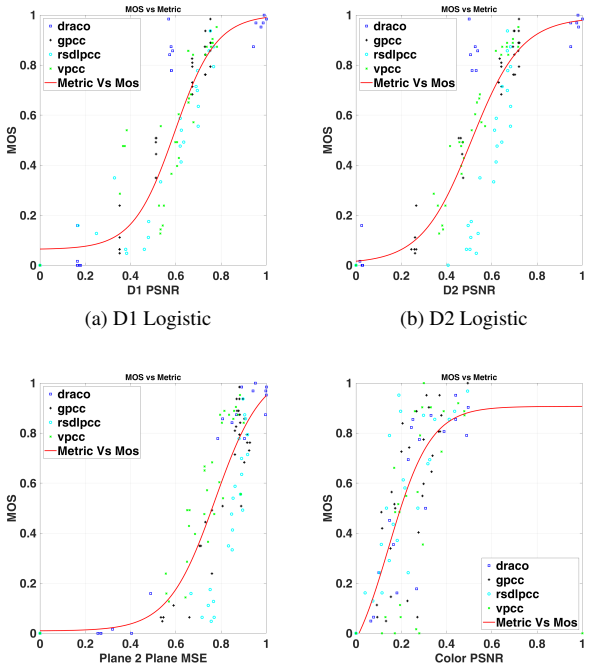
From table 5 and table 6, it can be observed that the best performing metrics were the point to point and point to plane metrics. A plot representing the relation between each metric and the MOS is represented in figure 7 including the logistic fitting curve. We can observe in this plot that while V-PCC and G-PCC data tends

Table 5: Linear Fitting

Metric	PCC	SROCC	RMSE	OR
po2point_MSE_PSNR	**0.862**	**0.884**	**0.163**	**0.735**
po2plane_MSE_PSNR	**0.814**	**0.847**	**0.187**	**0.784**
pl2plane_MSE	0.791	0.795	0.197	0.775
color_PSNR	0.488	0.679	0.280	0.833

Table 6: Logistic Fitting

Metric	PCC	SROCC	RMSE	OR
po2point_MSE_PSNR	**0.890**	**0.884**	**0.148**	**0.618**
po2plane_MSE_PSNR	**0.851**	**0.847**	**0.169**	**0.618**
pl2plane_MSE	0.846	0.795	0.172	0.667
color_PSNR	0.670	0.679	0.240	0.719



(a) D1 Logistic (b) D2 Logistic

(c) Plane 2 Plane Logistic (d) Color Logistic

Figure 7: Relation between metrics and MOS, and Logistic fitting curve.

to be very close from the logistic curve, RS-DLPCC tends to be below the curve. This reveals that the metrics provide a good representation for the codec, but they are not suitable to compare with other codecs. Furthermore, Draco results are not appropriately represented by these metrics. This is a case were metrics could misjudge the performance of a codec.

Conclusions

An evaluation on the quality of V-PCC, G-PCC, RS-DLPCC and Draco codecs is presented. This paper reveals that MPEG codecs are the best performing solutions. The tested deep learning solution also provided a very good compression performance result, with space to further improvements. It was developed for geometry compression and can still improve its performance through the appropriate compression of each point associated features (RGB components in this case). Moreover, better training and better architectures can be implemented. This deep learning

solution is likely to produce some blocking artifacts that we believe were the cause of the reduction of performance when compared with the MPEG codecs. Draco does not provide state of the art results for point cloud compression.

This research also revealed that the most common point cloud metrics fail to provide an accurate representation of quality when deep learning compression models are used. While the metrics show similar results for the G-PCC and the studied deep learning solution, the subjective evaluation revealed different quality. Because of that, correlations are slightly lower than 0.9 for the studied metrics. Although those are still acceptable results, they reveal that these metrics should be carefully considered when different compression technologies are used, causing different types of distortions. In the near future, a deeper analysis of the state of the art point cloud metrics will be conducted using this subjective quality evaluation data.

Another fact that was revealed is that sparse point clouds might tend to have different behaviors. In the case of bumbameuboi, which is rather sparse when compared with the others, G-PCC provided the most efficient representation and the deep learning solution provides a really bad performance.

Acknowledgemnts

This research was funded by the Portuguese FCT-Fundação para a Ciência e Tecnologia under the project UIDB/EEA/50008/2020, PLive X-0017-LX-20, and by operation Centro-01-0145-FEDER-000019 - C4 - Centro de Competencias em Cloud Computing.

The authors would like to acknowledge the authors of [20] for providing the deep learning decoded data for the reported experiment.

References

[1] JPEG Pleno Database, https://jpeg.org/plenodb/. [Online].

[2] UNIVERSITY OF SÃO PAULO POINT CLOUD DATASET http://uspaulopc.di.ubi.pt [Online]

[3] MPEG 3DG, "G-PCC Codec Description v5," ISO/IEC JTC1/SC29/WG11 N18891, Geneva, CH, October 2019.

[4] "V-PCC Test Model v8," ISO/IEC JTC1/SC29/WG11 W18884, Geneva, CH, October 2019.

[5] A. Guarda et al., "Deep Learning-based Point Cloud Geometry Coding with Resolution Scalability," IEEE International Workshop on Multimedia Signal Processing (MMSP'2020), Tampere, Finland, Sep 2020.

[6] A. Guarda, Nuno M. M. Rodrigues, F. Pereira, Adaptive Deep Learning-based Point Cloud Geometry Coding, IEEE Journal on Selected Topics in Signal Processing, Vol. 15, No. 2, pp. 415 - 430, February, 2021

[7] S. Perry et al., "Quality Evaluation Of Static Point Clouds Encoded Using MPEG Codecs," 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 3428-3432

[8] FFMPEG Documentation, https://ffmpeg.org/ffmpeg.html. [Online]

[9] E. Alexiou, E. Upenik, and T. Ebrahimi, "Towards subjective quality assessment of point cloud imaging in augmented reality," in 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), 2017, pp. 1–6.

[10] E. Alexiou et al., "Point cloud subjective evaluation methodology based on 2D rendering," in 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), 2018, pp. 1–6.

[11] A. Javaheri et al. "Subjective and objective quality evaluation of compressed point clouds" 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP) pp. 1-6 2017.

[12] L. A. da Silva Cruz et al., "Point cloud quality evaluation: Towards a definition for test conditions," in 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), 2019, pp. 1–6.

[13] S. Perry et al. "Study of subjective and objective quality evaluation of 3d point cloud data by the jpeg committee" Electronic Imaging vol. 2019 no. 10 pp. 12–1-312–7 2019

[14] E. Alexiou et al., "A comprehensive study of the rate- distortion performance in mpeg point cloud compression," APSIPA Transactions on Signal and Information Processing, vol. 8, p. e27, 2019.

[15] H. Su et al., "Perceptual quality assessment of 3d point clouds," in 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 3182–3186.

[16] D.Tian, H.Ochimizu, C.Feng, R.Cohen, and A.Vetro. Geometric distortion metrics for point cloud compression. In 2017 IEEE International Conference on Image Processing (ICIP), pages 3460–3464, 2017

[17] E. Alexiou, T. Ebrahimi, "Point cloud quality assessment metric based on angular similarity", IEEE International Conference on Multimedia and Expo (ICME), July 2018.

[18] CloudCompare: 3D point cloud and mesh processing software Open Source Project, https://www.danielgm.net/cc/, 2019.

[19] ITU-T P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," International Telecommunication Union, Jul. 2012.

[20] A. F. R. Guarda, N. M. M. Rodrigues and F. Pereira, "Point Cloud Coding: Adopting a Deep Learning-based Approach," 2019 Picture Coding Symposium (PCS), 2019, pp. 1-5, doi: 10.1109/PCS48520.2019.8954537.

[21] O. Devillers and P. . Gandoin, "Geometric compression for interactive transmission," in Proceedings Visualization, 2000, pp. 319–326.

[22] G. Bruder, F. Steinicke and A. Nüchter, "Poster: Immersive point cloud virtual environments," 2014 IEEE Symposium on 3D User Interfaces (3DUI), 2014, pp. 161-162, doi: 10.1109/3DUI.2014.6798870.

[23] Quach, Maurice and Valenzise, Giuseppe and Dufaux, Frederic, "Learning Convolutional Transforms for Lossy Point Cloud Geometry Compression", 2019 IEEE International Conference on Image Processing (ICIP),2019

[24] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union, Jan. 2012.

[25] BT709 ITU-R BT.709, Parameter values for the HDTV standards for production and international programme exchange, Jun. 2015.

[26] Google, Draco 3D Data Compression, https://github.com/google/draco. [online]

5.4 Subjective and Objective Testing in Support of the JPEG Pleno Point Cloud Compression Activity

Subjective and Objective Testing in Support of the JPEG Pleno Point Cloud Compression Activity

S. Perry, Luis A. Da Silva Cruz, J. Prazeres, António Pinheiro, Emil Dumic, Davi Lazzarotto, Touradj Ebrahimi

2022 10th European Workshop on Visual Information Processing (EUVIP), Lisbon, Portugal, 2022, pp. 1-6
DOI:10.1109/EUVIP53989.2022.9922803

Subjective and Objective Testing in Support of the JPEG Pleno Point Cloud Compression Activity

Stuart Perry
University of Technology Sydney
Sydney, Australia
Stuart.Perry@uts.edu.au

Luis A. da Silva Cruz
Instituto de Telecomunicações (IT)
Universidade de Coimbra
Coimbra, Portugal
lcruz@deec.uc.pt

João Prazeres, António Pinheiro
Instituto de Telecomunicações (IT)
Universidade da Beira Interior
Covilhã, Portugal
joao.prazeres@ubi.pt pinheiro@ubi.pt

Emil Dumic
Department of Electrical Engineering
University North, Croatia
Varazdin, Croatia
edumic@unin.hr

Davi Lazzarotto, Touradj Ebrahimi
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
davi.nachtigalllazzarotto@epfl.ch touradj.ebrahimi@epfl.ch

2022 10th European Workshop on Visual Information Processing (EUVIP) | 978-1-6654-6623-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/EUVIP53989.2022.9922803

***Abstract*—Point clouds have many applications in today's society ranging from entertainment to autonomous driving. With these new applications comes the need to compress the growing volume of point cloud data in a manner that is both suitable for human visualization and machine processing applications. The JPEG Pleno Point Cloud activity has been working toward a learning-based coding standard for point clouds, offering a single-stream, compact compressed domain representation, supporting advanced flexible data access functionalities targeting both interactive human visualization, and effective performance for 3D processing and machine-related computer vision tasks. As part of this activity, the JPEG Committee has been performing a number of exploration studies to evaluate existing coding standards as well set up baseline anchors and examine objective metrics against which new learning-based solutions may be compared. This article provides an overview of the JPEG Pleno Point Cloud activity and discusses challenges and solutions to the problem of evaluating and comparing cloud coding solutions. Experimental results will be presented demonstrating methodologies used by the JPEG Committee for point cloud compression assessment as well as outlining the performance of current state of the art compression standards on point clouds as well as the sensitivity of the objective metrics used for this activity to various adjustable parameters.**

***Index Terms*—JPEG, point cloud, compression, machine learning**

I. Introduction

Point cloud applications have become more numerous in the last 10 years and look to continue on an accelerated trajectory of adoption by society. Applications derived from 3D scanning, analysis and visualisation as well as augmented, mixed and virtual reality applications look to have a dramatic effect on society in the near future. These emerging applications create new challenges and demand new technologies to unlock their potential. One of the major emerging challenges is the massive volume of 3D data that needs to be collected, stored, analysed and displayed to enable the use of point clouds in practical applications. A high quality scan of even a small object can require millions of points to represent the object shape, while the unrestricted positions of the points in space together with the need to store attributes mean representation cost of the full point cloud can be large, easily reaching gigabytes. If one considers emerging applications such as autonomous driving that involve the capture and processing of streams of point cloud data in real time, the need for efficient and powerful compression technologies for point clouds becomes urgent. The JPEG Committee has been working on coding standards for plenoptic data as part of its JPEG Pleno activity for a number of years. Plenoptic data in this context is considered to cover holography, light fields and point clouds, all of which are different representations of the plenoptic capture function [1], [2]. The scope of the JPEG Pleno Point Cloud activity is the development of standards for point cloud representation that not only involve efficient coding, but also support machine vision applications. This activity will advance through a series of stages:

- Stage 1: A learning-based coding standard addressing human visualization and decompressed/reconstructed domain 3D processing and computer vision tasks;
- Stage 2: A learning-based coding standard additionally supporting compressed domain 3D processing such as visual enhancement and super-resolution;
- Stage 3: A learning-based coding standard additionally supporting compressed domain computer vision tasks such as classification, recognition and segmentation.

During the 94th JPEG meeting, the JPEG Committee released

António M. G Pinheiro and João Prazeres thank the Portuguese FCT-Fundação para a Ciência e Tecnologia under the project UIDB/EEA/50008/2020, PLive X-0017-LX-20, and operation Centro-01-0145-FEDER-000019 - C4 - Centro de Competencias em Cloud Computing for funding this research.
Davi Lazzarotto and Touradj Ebrahimi thank the Swiss National Foundation for Scientific Research (SNSF) under the grant number 200021-178854 for funding this research.

a Final Call for Proposals on JPEG Pleno Point Cloud Coding. This call addresses Stage 1 of the activity [3]. In early 2022 a study was performed to support the Call for Proposals. The goal of this study was to:

1) Evaluate the performance of current state of the art point clouds coding solutions tested on samples of the point cloud training set to be supplied to proponents for the Call for Proposals.
2) Understand to which extent differences between laboratories may affect the subjective quality assessment of submissions to the upcoming Call for Proposals.
3) Determine the impact of point cloud normal estimation methods and parameters on the computation of objective metrics intended to be used during the Call for Proposals.

In Section II the experimental methodology followed will be presented including the selection of point clouds and state of the art point cloud codecs for use in the study, as well as the objective metrics and the subjective testing methodology to be used. Section III will detail the results of the study, while Section IV will provide a discussion of the results.

II. Experimental Setup

A. Content

To benchmark the performance of the chosen state of the art codecs, a set of seven point clouds were chosen. The point clouds used in this investigation are shown in Fig. II-A. The *longdress*, *guanyin* and *rhetorician* point clouds were sourced from the JPEG Pleno Database [4]. The *camera*, *car*, *plantanopote* and *suzuki* point clouds are sampled from meshes obtained from the ShapeNetCore Database [5]. The dataset is publicly available[1]. The sampling process followed Lazzarotto and Ebrahimi's methodology [6], which involved the exclusion of internal faces prior to the sampling process in order to avoid obtaining colors from different faces at similar positions.

B. Anchor Codecs

In order to establish a baseline for the future comparison of learning-based point cloud codecs, the JPEG Committee chose two common non-learning-based codecs developed by the MPEG Standardisation group; G-PCC [7] and V-PCC , [8], [9] as anchor codecs. These codecs will form the base level of performance for the subsequent Call for Proposals on JPEG Pleno Point Cloud Coding [3], so it is imperative that the performance on the training set is well understood.

G-PCC uses an octree encoding method. It has two encoding modes for the deepest level of geometrical information; Octree and Triangle Soup. In this work the Octree encoding mode was selected, with compression factor controlled by the *positionQuantizationScale* parameter to obtain five encoding rates (R01-R05) from low to high quality. For each of the rates, the Lifting parameters *seq_lod* and *seq_dist2* were set to 12 and 3 respectively.

[1]http://webx.ubi.pt/∼pinheiro/euvip2022pcdb.html

Fig. 1. Point Clouds used in this investigation. The *longdress*, *guanyin* and *rhetorician* point clouds were sourced from the JPEG Pleno Database [4], while *camera*, *car*, *plantanopote* and *suzuki* point clouds are sampled from meshes obtained from the ShapeNetCore Database [5] using the technique of Lazzarotto and Ebrahimi [6].

V-PCC uses a projection based method wherein the point cloud is projected as a set of patches onto multiple planes (usually six). The projection patches represent point cloud texture and color, depth information and an occupancy map. Each projected set of patches is compacted and the resulting sequence of images compressed using traditional 2D video techniques. MPEG V-PCC test model TMC2 version 8 [9] with VVC was used in All Intra (AI) coding mode with the encoding condition being *C2, Lossy Geometry - Lossy Attributes*.

C. Objective Metrics

Currently, there is already a wide variety of point cloud quality metrics available. Based on a previous study [10], the JPEG Committee has found that the PSNR D1 and PSNR D2 [11] quality metrics display consistent performance in terms of point cloud quality evaluation. Since PSNR D1 and PSNR D2 only measure geometrical accuracy of point clouds, there is a need to include additional metrics that use both color and geometry. For the purpose of supporting the Call for Proposals, the authors considered some recent point cloud quality measures, PCQM [12] and PointSSIM [13] [14]. For each point cloud/codec/rate combination, the objective quality metrics PSNR D1, PSNR D2, PCQM and PointSSIM were computed. The PSNR D1 and PCQM measure point to point distances, whereas PSNR D2 requires normal information to measure surface to point distances and PointSSIM can also be employed with normal-based features or color-based features. To compute the PointSSIM metric, the variance (VAR) was used as a statistical estimator, and a neighborhood size of 12 was used as recommended in the original work [13]. Both normal-based and color-based features were considered. The use of normal information in objective metric calculations can lead to inconsistent results, particularly for sparse point clouds. Depending on spatial sparsity of the points and the method of normal calculation, the obtained metrics can be subject to unwanted variation. In this work, an investigation was conducted to inquire whether the number of neighbouring points used to compute the normals had an effect on the accuracy of PSNR D2 and PointSSIM. To do so, Cloud Compare [15] was used to fit a quadric local surface based on 5, 10 and 20 neighbour points from which the normals were computed. The estimated normals were then used to compute PSNR D2 and PointSSIM values.

D. Subjective Testing Methodology

For the subjective quality component of this experiment, a set of 12 second stimulus videos at 4096x2160 resolution were created with reference and processed (encoded by a codec at a particular rate point and then decoded to create a reconstruction) point clouds shown side by side. The videos were shown to subjects at a frame rate of 30fps using a customised version of the MPV video player [16]. During the 12 second period, the reference and processed point clouds were rotated synchronously about their respective central vertical axes, to complete a full 360° path. Subjects were instructed to judge visual quality of the processed with respect to the reference point clouds according to a Double Stimulus Impairment Scale protocol with 5 possible impairment ratings (1 - *very annoying*, 2 - *annoying*, 3 - *slightly annoying*, 4 - *perceptible, but not annoying* and 5 - *imperceptible*). To mitigate potential bias, each subject was only shown videos with the reference on the same side of the display, with half of subjects shown videos with the reference on the left and the remaining half of the subjects shown videos with the reference on the right. The content presentation order was random, but adjusted so that subjects did not at any point see the same content as that shown in the immediately preceding video. Each session started with a training session using a point cloud from the JPEG Pleno Database [4] that was not used for subsequent data collection. Following the training session, subjects were shown seven different content types processed by two codecs at five different rates together with seven reference-reference pairs (one for each content point cloud) for a total of 77 double stimuli videos. The reference-reference pairs were included to understand subject behaviour in the case when no artefacts were present and to determine if non-attentive subjects were present. Data from two test labs is described in this work: University of Beira Interior (UBI), Covilhã, Portugal and University North (UNIN), Varaždin, Croatia and test environments were set up according to ITU-R Recommendation BT.500-13 [17] as shown in Table I. The display resolution used by UNIN is smaller than the videos, but as no video scaling was allowed, the videos were displayed in true resolution. After careful check, was observed that the information of the point cloud was not cropped. This means that subjects in UBI and UNIN saw the same information. The cropped area did not show any point cloud information in all cases. Outlier detection was performed according to BT.500-13 [17] on each laboratory set of data separately with no outliers found. Finally mean opinion scores (MOS) and 95% confidence intervals were computed. Table II presents the gender and age breakdowns of the subjects for the two laboratories.

TABLE I
EXPERIMENTAL SETUP AT TEST LABORATORIES

Laboratory	Display Type	Resolution	Viewing Distance
UBI	Eizo ColorEdge CG318-4K	4096x2160 (31.1″)	1.2m (FV ±15cm)
UNIN	Sony TV 55” KD-55x8505C	3840x2160 (55″)	1.5m (FV ±15cm)

TABLE II
TEST SUBJECT INFORMATION AT TEST LABORATORIES

	Males	Females	Total	Age span	Average age
UBI	10	8	18	21-34	26.0
UNIN	17	1	18	19-59	26.4

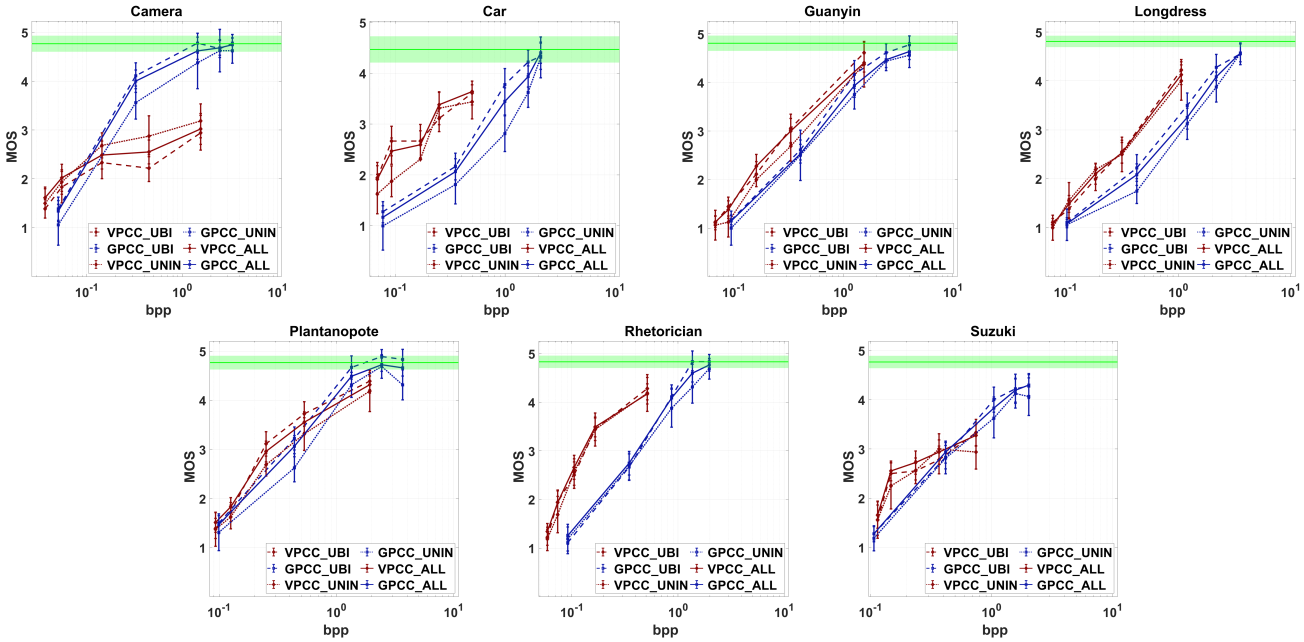


Fig. 2. MOS results for the seven tested point clouds. The red lines represent results for the V-PCC codec, while the blue lines represent results for the G-PCC codec. The error bars are 95% confidence intervals, while the green bars represent the 95% confidence intervals for the reference-reference stimuli.

TABLE III
CORRELATION OF MOS RESULTS ACROSS LABS

	PCC	SROCC	RMSE	OR
UBI vs UNIN	0.983	0.979	0.062	0.143

III. RESULTS

A. Subjective Results

Figure 2 shows the MOS plotted against bitrate for individual labs and aggregated across all the subjects from all of the participating laboratories. Bitrate is measured as bits per point (bpp) and is computed as the ratio of the total number of bits of the encoded content divided by the number of input points in the encoded point cloud. Based on the high degree of correlation found between the different laboratories, as will be demonstrated in Section III-B, the authors considered the consolidation of the scores from all the laboratories to be valid.

B. Correlation Across Labs

To determine the degree of correspondence of MOS between the different test laboratories, the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), Root-Mean Squared Error (RMSE) and Outlier Ratio (OR) were computed. The results are presented in Table III. Figure 3 shows the linear fitting across all laboratories. In general, the correlation between the test laboratories is quite high with both Pearson and Spearman correlation coefficients above 0.97.

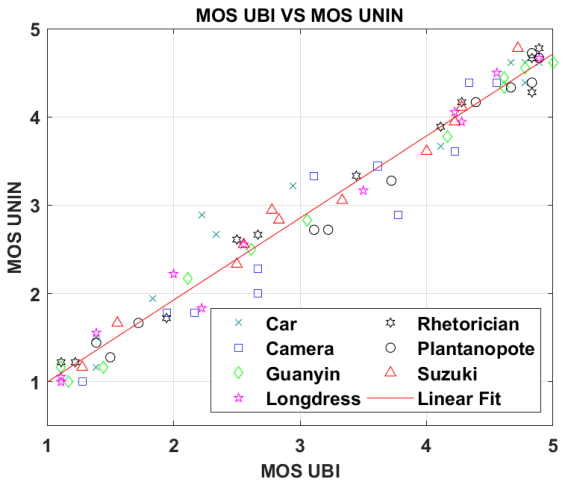


Fig. 3. Linear fitting for correlation between MOS obtained from different test laboratories.

C. Objective Metric Results

To measure the ability of the objective metrics to predict subjective scores, we employed the methodology from Recommendation ITU-T P.1401 [18]. This involves the computation of PCC, SROCC, RMSE and OR on the original and predicted MOS values. The predicted MOS values were obtained following the fitting of a logistic function to the objective scores. The results are shown in Table IV, while the individual MOS-objective quality pairs and fitted curves are shown in Fig. 4.

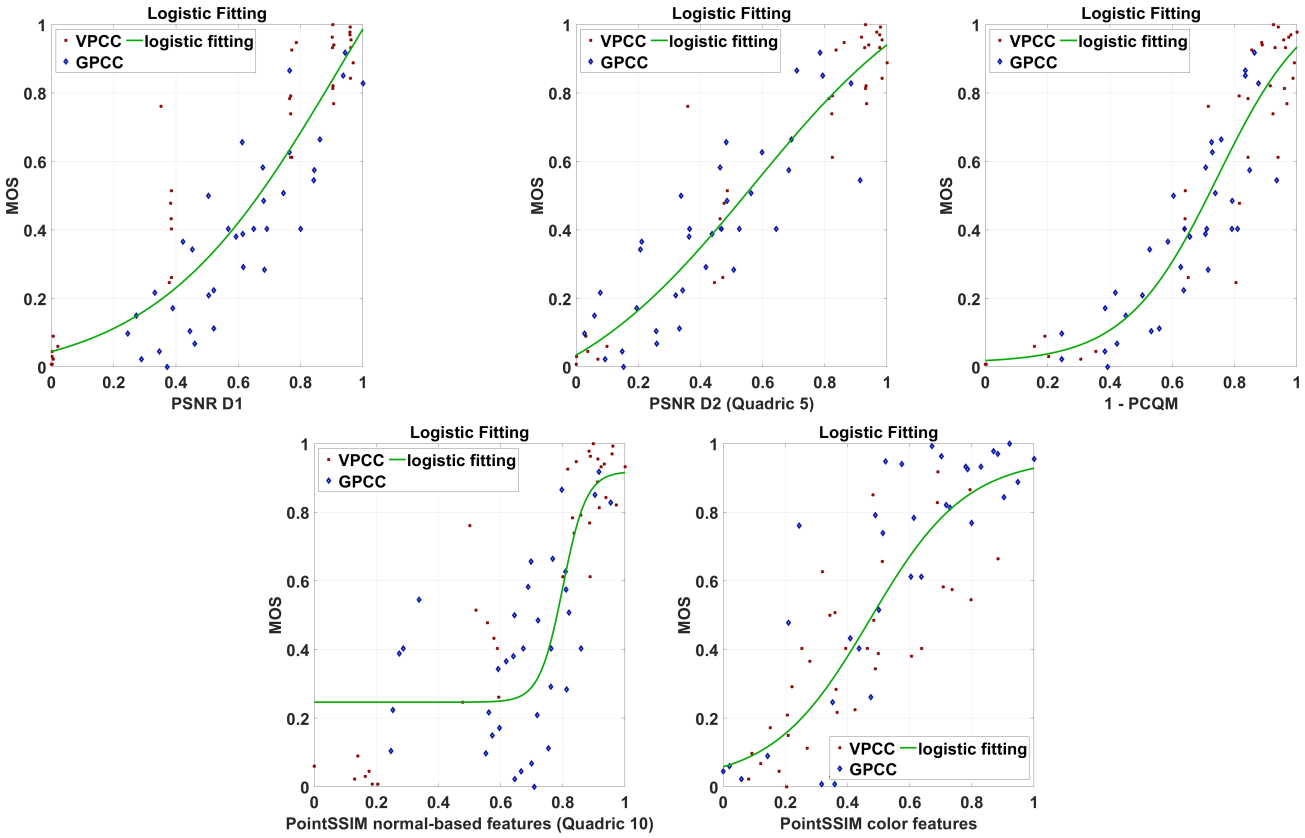


Fig. 4. MOS results plotted against objective metrics and regression curve between metric and MOS.

TABLE IV
PERFORMANCE MEASURES OF OBJECTIVE METRICS AGAINST MOS RESULTS.

Metric	PCC	SROCC	RMSE	OR
PSNR D1	0.889	0.879	0.151	0.529
PSNR D2 Quadric 5	0.928	0.921	0.123	0.457
PSNR D2 Quadric 10	0.928	0.918	0.123	0.443
PSNR D2 Quadric 20	0.926	0.916	0.125	0.443
PointSSIM Quadric 5	0.761	0.765	0.213	0.671
PointSSIM Quadric 10	0.824	0.800	0.186	0.700
PointSSIM Quadric 20	0.765	0.763	0.212	0.686
PointSSIM Color-based	0.830	0.827	0.184	0.600
PCQM	0.916	0.913	0.132	0.586

IV. DISCUSSION AND CONCLUSIONS

Based on the results described in Section III, a number of conclusions can be drawn. The results from the two laboratories were highly consistent despite the use of different displays and different resolutions. This robustness has been observed in previous studies [10], [19] and is encouraging as this is crucial to the ability of the JPEG Committee to accurately ascertain the performance of proposals. Examining Fig. 2 it can be observed that although different laboratories have similar MOS for the same content, there are clear differences between the performance G-PCC and V-PCC dependent on content. For example, for most of the content, V-PCC outperforms or performs as well as G-PCC with the exception of the *camera* point cloud where at higher bitrates G-PCC performs better. It is unclear as to what aspects of the *camera* point cloud might be responsible. The point cloud has a number of large flat surfaces that may have been difficult for V-PCC to encode accurately when not aligned precisely with the projection surfaces. For the objective metric results, we observe in Fig. 4 that PSNR D1, PSNR D2 and PCQM show a good relationship between the objective metrics and MOS, however PointSSIM appears to have reduced accuracy in MOS prediction at lower quality levels for the version that makes use of normal features. Higher compression levels for point clouds are generally associated with an increased sparsity of the reconstructed point cloud. The reduced accuracy of the PointSSIM metric may be related to the increased sparsity of

the point clouds at lower quality levels. From Table IV we can observe a decreased correlation and an increased Outlier Ratio for PointSSIM compared to the other metrics. In regard to the effect that normal calculation has on the accuracy of the metrics, we can observe that while PSNR D2 appears to be relatively robust to the number of neighbouring points used in the normal calculation, PointSSIM appears more sensitive to this factor. Pearson correlation values range from 0.761 to 0.824 with the addition of normal information, below the value of 0.830 when color-based features are employed. This increased sensitivity of PointSSIM to the normal vectors is probably due to the fact that the estimation has to be performed for both the reference and the degraded models, contrary to PSNR D2 which only requires normal vectors for the reference.

ACKNOWLEDGMENT

The authors thank Nhung Thi Hong Nguyen for assistance with the preparation of stimuli for the experiment. We also thank University of Beira Interior (PT), Vrije Universiteit Brussel (BE), University North (HR) and University of Patras (GR) for participating in the subjective experiment. Unfortunately, due to low numbers or differences in display resolution data from Vrije Universiteit Brussel (BE) and University of Patras (GR) could not included in this work.

REFERENCES

[1] P. Schelkens, Z.Y. Alpaslan, T. Ebrahimi, K.-J. Oh, F.M.B. Pereira, A.M.G. Pinheiro, I. Tabus, Z. Chen, "JPEG Pleno: a standard framework for representing and signaling plenoptic modalities," Proc. SPIE 10752, Applications of Digital Image Processing XLI, 107521P (17 September 2018)

[2] P. Astola, L. da Silva Cruz, E. da Silva, T. Ebrahimi, P. Freitas, A. Gilles, K. Oh, C. Pagliari, F. Pereira, C. Perra, S. Perry, A. Pinheiro, P. Schelkens, I. Seidel, I. Tabus, "JPEG Pleno: Standardizing a coding framework and tools for plenoptic imaging modalities", ITU Journal: ICT Discoveries, vol. 3, no. 1, June 2020.

[3] ISO/IEC JTC1/SC29/WG1, "Final Call for Proposals on JPEG Pleno Point Cloud Coding," Doc. WG1N100097, Jan 2022.

[4] JPEG Pleno Database, https://jpeg.org/plenodb/. [Online]. Available: https://jpeg.org/plenodb/.

[5] ShapeNet, https://shapenet.org/. [Online]. Available: https://shapenet.org/.

[6] D. Lazzarotto and T. Ebrahimi, "Sampling color and geometry point clouds from ShapeNet dataset", arXiv:2201.06935, Jan 2022.

[7] MPEG 3DG, "G-PCC Codec Description v5," ISO/IEC JTC1/SC29/WG11 N18891, Geneva, CH, October 2019.

[8] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "Emerging mpeg standards for point cloud compression," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 133–148, March 2019.

[9] MPEG 3DG, "V-PCC Test Model v8," ISO/IEC JTC1/SC29/WG11 W18884, Geneva, CH, October 2019.

[10] S. Perry, H.P. Cong, L.A. da Silva Cruz, J. Prazeres, M. Pereira, A. Pinheiro, E. Dumic, E. Alexiou, T. Ebrahimi, "Quality Evaluation of Static Point Clouds Encodec Using MPEG Codecs," Proceedings of the 27th IEEE International Conference on Image Processing 2020 (ICIP2020), Abu Dhabi, United Arab Emirates, 25-28 October 2020.

[11] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Evaluation metrics for point cloud compression," ISO/IEC JTC m74008, Geneva, Switzerland, Tech. Rep., January 2017.

[12] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "PCQM: a full-reference quality metric for colored 3d point clouds," in the Proceedings of the 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6.

[13] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in the Proceedings of the 2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2020, pp. 1–6.

[14] D. Lazzarotto; E. Alexiou; T. Ebrahimi, "Benchmarking of objective quality metrics for point cloud compression", in the Proceedings of the 2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP), 2021, pp. 1-6.

[15] Cloud Compare, https://cloudcompare.org/doc/wiki/. [Online]. Available: https://cloudcompare.org/doc/wiki/.

[16] MPV video player, https://mpv.io. [Online]. Available: https://mpv.io.

[17] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union, Jan. 2012.

[18] ITU-T P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," International Telecommunication Union, Jul. 2012.

[19] S. Perry, L.A. da Silva Cruz, E. Dumic, N.H.T. Nguyen; A. Pinheiro, E. Alexiou, "Comparison of Remote Subjective Assessment Strategies in the Context of the JPEG Pleno Point Cloud Activity", Proceedings of the 2021 IEEE International Workshop on Multimedia Signal Processing (MMSP 2021), Tampere, Finland, 6-8 October 2021.

5.5 On the stability of point cloud machine learning based coding

On the stability of point cloud machine learning based coding
J. Prazeres, R. Rodrigues, M. Pereira and A. M. G. Pinheiro
2022 10th European Workshop on Visual Information Processing (EUVIP), Lisbon, Portugal, 2022, pp. 1-6
DOI: 10.1109/EUVIP53989.2022.9922676

On the stability of point cloud machine learning based coding

João Prazeres, Rafael Rodrigues, Manuela Pereira, and Antonio M. G. Pinheiro
Instituto de Telecomunicações & Universidade da Beira Interior, Covilhã, Portugal
joao.prazeres@ubi.pt, rafael.rodrigues@ubi.pt, mpereira@di.ubi.pt, pinheiro@ubi.pt

2022 10th European Workshop on Visual Information Processing (EUVIP) | 978-1-6654-6623-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/EUVIP53989.2022.9922676

***Abstract*—This paper analyses the performance of two of the most well known deep learning-based point cloud coding solutions, considering the training conditions. Several works have recently been published on point cloud machine learning-based coding, following the recent tendency on image coding. These codecs are typically seen as a set of predefined trained machines. However, the performance of such models is usually very dependent of their training, and little work has been considered on the stability of the codecs' performance, as well as the possible influence of the loss function parameters, and the increasing number of training epochs. The evaluation experiments are supported in a generic test set with point clouds representing objects and also more complex scenes, using the point to point metric (PSNR D1), as several studies revealed the good quality representation of this geometry-only point cloud metric.**

***Index Terms*—Point cloud coding, machine learning-based codecs**

I. Introduction

The usage of 3D data formats has been increasing recently, most notably in virtual (VR), augmented (AR) and mixed reality applications, but also in a wide variety of other fields, such as computer graphics, 3D printing, construction, manufacturing, robotics, automation, medical applications, retail, cultural heritage, remote sensing, and geographical information systems.

Point cloud technology is one of the most popular solutions for 3D data representation, which maps surfaces on a Cartesian coordinate system (x, y, z). Each mapped point might have a list of associated attributes, including RGB components, reflectance, physical sensor information, or normal vectors. Point clouds can provide accurate representations of both objects and scenes, from any viewing position or distance.

The models of data representation and their associated quality play an essential role in point cloud applicability, as 3D content often creates huge amounts of information. Accurate point cloud representations of landscapes, buildings or artefacts typically contain several millions of points, each with one or more associated attributes. Thus, depending on the represented object or scene, the size of raw point clouds may become impractical, especially in real-time applications. For that reason, efficient point cloud compression and decompression solutions are needed.

Research funded by the Portuguese FCT-Fundação para a Ciência e Tecnologia under the project UIDB/50008/2020, PLive X-0017-LX-20, and by operation Centro-01-0145-FEDER-000019 - C4 - Centro de Competencias em Cloud Computing.

Two of the most recognized and used coding solutions were developed by MPEG, i.e., the Video-based Point Cloud Compression (V-PCC) [1] and the Geometry-based Point Cloud Compression (G-PCC) [2]. Authors in [3] provide a quality study of both codecs. Google developed DRACO[1], which provides both lossless and lossy point cloud encoding. As shown in [4], the MPEG codecs perform better than DRACO, in terms of quality vs bit rate.

Recently, machine learning-based point cloud coding solutions have been emerging, following this trend on image and video coding. The Multiscale Point Cloud Geometry Compression (PCGC) was proposed by Wang *et al.* [5], and further developed in [6]. In [7] the Deep Point Cloud Geometry Compression (PCC_GEO_CNN) was presented, with an improved version proposed in [8]. Adaptive Deep Learning Point Cloud Compression was presented in [9]. In [10], a point cloud lossy attribute auto encoder is proposed, directly encoding and decoding attributes with the help of geometry. In [11], a deep convolutional autoencoder is proposed that directly operating on the points. It also considers a deconvolution operator in order to upsample point clouds, allowing decompression to an arbitrary density.

When dealing with learning-based methods, it is well known that the performance of the final learned model may vary, even with similar training conditions, due to the stochastic nature of the learning process. However, most research efforts in deep learning-based coding report their results as unique. In this work, the performance of two learning-based codecs is assessed based on the robustness of the compression performance from different training processes, under similar conditions. More specifically, the stability of PCGC [6] and PCC_GEO_CNN [8] is assessed, based on the resulting coding performances. For each step of the learning progression, a set of six point clouds is encoded and the point to point metric (PSNR D1) [12] is computed. The PSNR D1 metric was chosen, as the codecs only encode geometry. The obtained results are analysed to verify if there is a convergence to a stable operating point. The point cloud point to point metric (PSNR D1) is used in this work as it revealed to be always one of the best that uses the geometry only [3], [4].

[1]https://github.com/google/draco

II. Codecs Architecture description

A. Multiscale Point Cloud Geometry Compression

The Multiscale Point Cloud Geometry Compression (PCGC) model [6] features a mirrored encoder-decoder architecture, which performs consecutive downsampling to multiple scales. The encoder consists of three sparse convolution modules in sequence, each containing two convolutional layers with 3x3x3 kernels and ReLU activations. In the second convolution layer, a stride of 2x2x2 is used for downsampling. After the sparse convolution modules, a residual feature extraction step is added, consisting of three instances of the Inception Residual Network [13]. The final latent representation Y is given by an additional sparse convolution layer, with 8 3x3x3 filters.

At the bottleneck, the two components of Y, i.e., the geometry coordinates (C_Y) and feature attributes (F_Y) are encoded separately, using the lossless octree codec [2] and entropy coding, respectively. In the decoding process, C_Y and F_Y are merged and upsampled through a convolutional branch that mirrors the encoder. In each module, transposed convolutions with stride 2x2x2 are used to upsample the sparse point clouds. After each upsampling step, a single 3x3x3 convolution with sigmoid activation is used to obtain the voxel occupation probability (p_i), which is in turn used to compute a scale-specific binary cross-entropy (BCE) loss:

$$L_{BCE} = -\frac{1}{N}\sum_{i}^{N}\left(y_i \log(p_i) + (1-y_i)\log(1-p_i)\right) \quad (1)$$

where y_i is i^{th}-voxel true label (1 if occupied, 0 otherwise).

The model training aims at optimizing the following Lagrangian loss function:

$$L = R + \lambda D \quad (2)$$

where D refers to the geometrical distortion and R is the resulting bit rate of feature attribute encoding ($\hat{F}_y$). D is obtained from the multi-scale BCE loss, which is the average of the BCE losses computed at each scale. Finally, the λ parameter sets the rate-distortion trade-off. Each time the λ value decreases, the training model will define a new working point where the bit rate is lower but the distortion measure will also decrease, as it has lost weight in the loss function.

B. Deep Point Cloud Geometry Compression

The Deep Point Cloud Geometry Compression (PCC_GEO_CNN) model [7] has a fairly straightforward learning-based approach, which proposes to reduce the blocking effect typically introduced by other learning-based codecs.

The model architecture features an encoder part (f_a) with three sequential convolutional layers, each with 32 filters. The first layer uses a 9x9x9 kernel with stride 2x2x2, whereas the other two use a 5x5x5 kernel with stride 2x2x2. A ReLU activation is used in the first two layers. The latent representation $y = f_a(x)$ is given by the linear output of the third layer. This latent representation is then quantized ($\hat{y} = Q(y)$) through element-wise integer rounding. $\hat{y}$ compression is performed using the Deflate algorithm, which is a combination of LZ77 and Huffman coding [14].

The decoder branch f_s consists of three transposed convolutional layers, which mirror the encoder in terms of number of filters (except for the last layer), kernel size and stride. All layers in the decoder use a ReLU activation function. The last layer uses a single filter and provides the distorted point cloud $\tilde{x}$, using element-wise minimum, maximum and rounding functions. In the decoding process, p_z^t is the probability of a point z being occupied. The global loss function of the PCC_GEO_CNN model is the same as the Eq. 2. The distortion component is given by the overall of the focal loss (Eq. 3),

$$L_{Focal}(x,\tilde{x}) = \alpha_z(1-p_z^t)^{\gamma} log(p_z^t) \quad (3)$$

which allows to compensate for the imbalanced ratio between occupied and non-occupied voxels. Given the balance parameter α, α_z is 1 if the corresponding voxel in the original point cloud is occupied, and 1-α otherwise.

III. Experimental setup and results

The experiments carried out will study the evolution of the codecs through the learning process. For that, the test data set will be coded/decoded after each epoch and the PSNR D1 is computed. The process is repeated three independent times to observe the stability of the codecs. The complete test will allow to understand the stability of the codecs with the same training conditions.

A. Test data selection

In this work, the performance of the tested point cloud codecs was assessed on a test set comprised of six point clouds, available at the JPEG Pleno database[2]. Three of these point cloud depict objects and the other three depict landscapes. The chosen object point clouds were the *Romanoillamp*, from the University of Sao Paulo Database[3], the *Guanyin* from the EPFL dataset, and frame 1300 of the *Longdress* dynamic point cloud. The selected landscapes are three point clouds from the University of Sao Paulo Database, namely *Citiusp*, *IpanemaCut* and *Ramos*. The six selected point clouds are shown in Figure 1.

B. PCGC Model Training

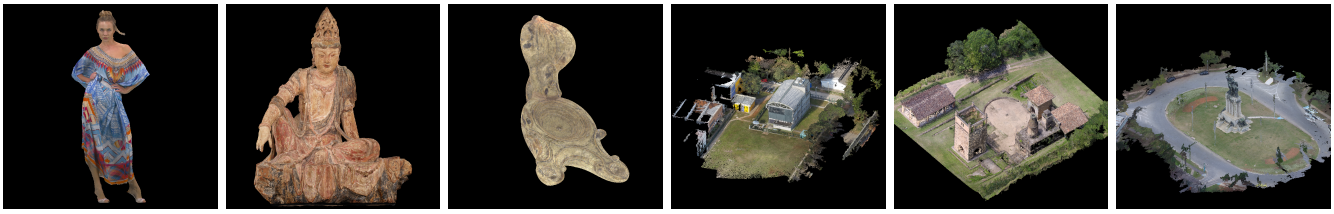
A study of the influence of model training on the performance of the PCGC codec [6] is established. This model was selected because information is given on the training procedures and even the training datasets are shared[4].

In [5], different coding bit rates are targeted, by varying the rate-distortion trade-off parameter λ between 0.75 and 16. In the code made available, the global loss function J depends on two parameters, α and β, such that $J = \alpha D + \beta R$. In this experiment, β was fixed at 1, so that α becomes equivalent to λ in Eq. 2.

[2]http://plenodb.jpeg.org/pc/8ilabs

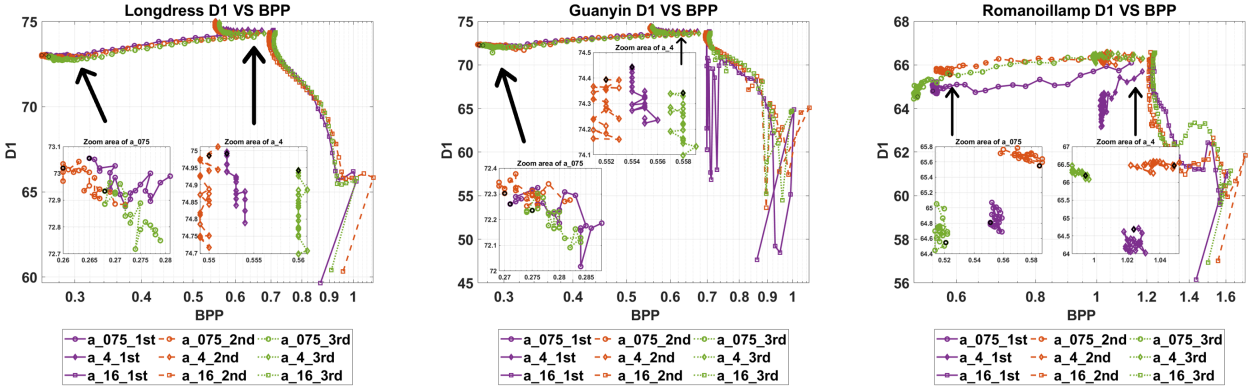
[3]http://uspaulopc.di.ubi.pt

[4]available at https://github.com/NJUVISION/PCGCv2

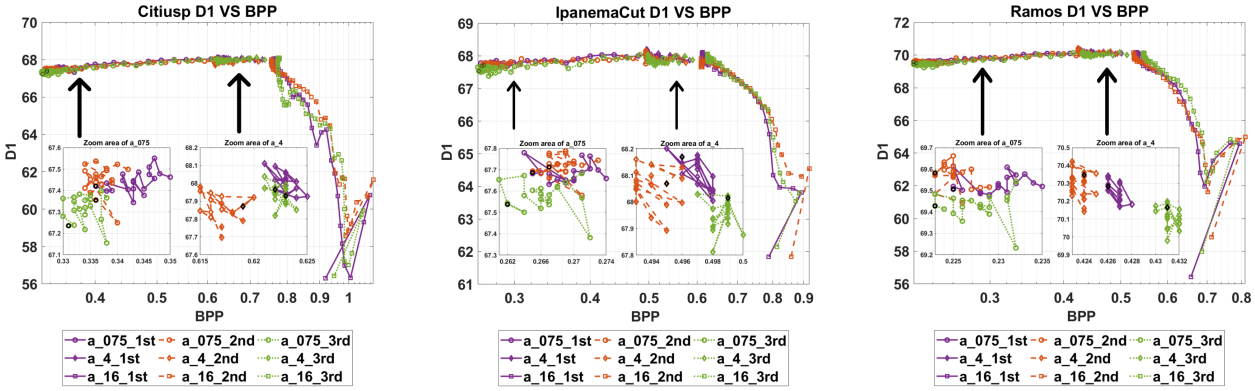


(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Fig. 1: Point Cloud test set.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp*

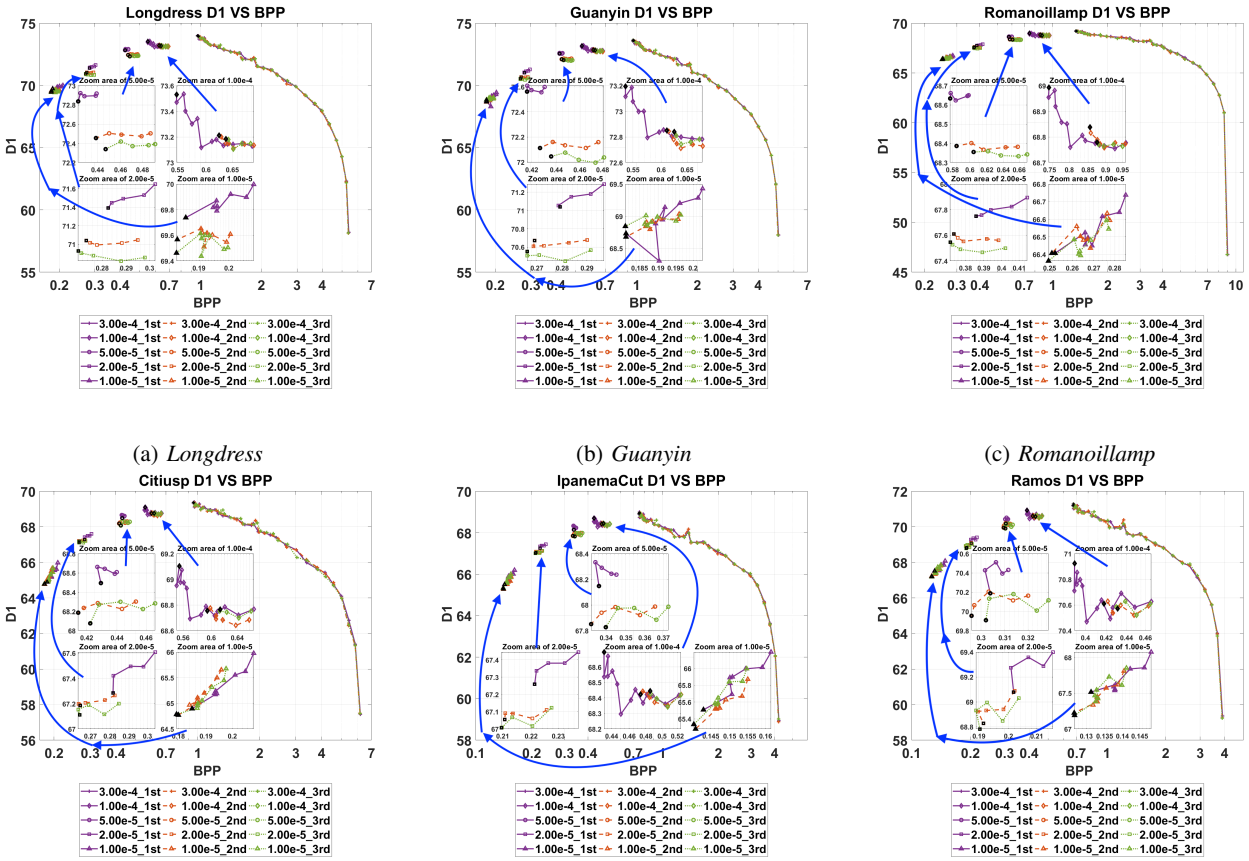


(d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Fig. 2: PSNR D1 vs. bpp plots for PCGC, trained with $\alpha = \{16, 4, 0.75\}$.

The model was trained with densely sampled data from the ShapeNet [15], a database containing $\approx$ 51.300 3D surface models. The final training set was obtained by random rotation and quantization with 7-bit precision. Also, the number of points in each point cloud was also randomized. In this paper, PCGC was trained with $\alpha = \{16, 4, 0.75\}$, each for 50 epochs, with a constant learning rate of 10^{-5}. For faster convergence, the learned weights with $\alpha = 16$ were used to initialize the training with both $\alpha = 4$ and $\alpha = 0.75$, as recommended in [5]. This training routine was ran three times with similar conditions. The result of the training sessions is shown in figure 2, with zoomed areas of interest.

The PCGC codec shows an acceptable level of stability for the *Citiusp*, *Longdress*, *IpanemaCut* and *Ramos*. In all of these cases, the D1 metric shows a similar behaviour across the training process. For the *Citusp* case, when $\alpha = 16$, some instability is shown between different epochs, across the different training sessions. For the point cloud *Guanyin*, a high level of instability is observed when $\alpha = 16$ for the first training session. In this case, D1 shows an inconsistent behaviour across the first training session, while the second and third sessions, a different, more stable behavior is observed,



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp*

(d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Fig. 3: PSNR D1 vs. bpp plots for PCC_GEO_CNN, trained with $\lambda = \{3 \times 10^{-4}, 10^{-4}, 5 \times 10^{-5}, 2 \times 10^{-5}, 10^{-5}\}$.

although some instability is found in higher bit rates. This is not observed for $\alpha = 4, 0.75$, where all the training sessions show very little D1 variation. In the encoding process of the *Romanoillamp* point cloud, the codec shows a high amount of instability. The D1 metric shows inconsistent behaviour across different training sessions. Contrary to what is observed for the other contents, the bit rate does not converge to a stable operating point, except for $\alpha = 16$.

In most cases, the codec reveals a good level on the encoding performance stability. Across all training sessions epochs, the bit rates converge to similar operating points. However, some content might show some undesirable change in the encoding performance, depending on the training process. Moreover, there is the assumption that none of the point clouds used in this test is present in the training data as they are not included in the Shape Net database, up to the authors knowledge.

C. PCC_GEO_CNN model training

In [8], the authors train four individual models for each Rate-Distortion tradeoff. They chose four values for λ (eq. 2), notably 3×10^{-4}, 10^{-4}, 5×10^{-5}, 2×10^{-5}). In the software provided by the authors[5], an additional value is considered, $\lambda = 10^{-5}$. This experiment followed the sequential training approach described in [8], where each training using λ_i is initialized with the trained weights of the previous model (i.e., using λ_{i-1}). The first training uses $\lambda_1 = 3 \times 10^{-4}$, and targets a low distortion, high bit rate model. Subsequent training uses λ values in descending order, which results in a progressive reduction of the target bit rates, while attempting to minimize the increase in distortion. The α and γ parameters of the focal loss function were set to 0.9 and 2, respectively, which were the default values in the provided code.

The test point clouds were encoded at each 500 training steps, which is the validation interval defined in the provided code. In the original code, the model checkpoint was saved at a given validation point, only if there is an improvement in the loss from the last validation point. However, for this experiment, the code was adapted to bypass this definition and save the checkpoints at every 500 training steps, to encode the test point clouds. An early stopping condition was also implemented in the original code, which interrupts the training

[5]avaliable at https://github.com/mauriceqch/pcc_geo_cnn_v2

(a) *Guanyin* cropped area of the original

(b) 1st training session, epoch=50, $\alpha = 0.75$, PSNR D1=64.808

(c) 2nd training session, epoch=50, $\alpha = 0.75$, PSNR D1=65.55

(d) 3rd training session, epoch=50, $\alpha = 0.75$, PSNR D1=64.5461

(e) 1st training session, step 31500, $\lambda = 10^{-5}$, PSNR D1=66.4068

(f) 2nd training session, step 28000,$\lambda = 10^{-5}$, PSNR D1=66.4068

(g) 3rd training session, step 27000,$\lambda = 10^{-5}$, PSNR D1=66.3627

Fig. 4: Decompressed *Guanyin* (Cropped area) for the low bit rate (final epoch) of each codec training. The first row shows the decoded point clouds for PCGC, and the second row for PCC_GEO_CNN.

(a) *Romanoillamp* cropped area of the original

(b) 1st training session, epoch=50, $\alpha = 0.75$, PSNR D1=64.808

(c) 2nd training session, epoch=50, $\alpha = 0.75$, PSNR D1=65.55

(d) 3rd training session, epoch=50, $\alpha = 0.75$, PSNR D1=64.5461

(e) 1st training session, step 31500, $\lambda = 10^{-5}$, PSNR D1=66.4068

(f) 2nd training session, step 28000, $\lambda = 10^{-5}$, PSNR D1=66.4068

(g) 3rd training session, step 27000, $\lambda = 10^{-5}$, PSNR D1=66.3627

Fig. 5: Decompressed *Romanoillamp* (Cropped area) for the low bit rate (final epoch) of each codec training. The first row shows the decoded point clouds for PCGC, and the second row for PCC_GEO_CNN.

process if the loss does not improve for more than 4 validation steps.

The models are trained on a subset of the ModelNet40 [16] dataset. First, the mesh data is voxelized with resolution $512 \times 512 \times 512$ and the 200 largest point clouds are selected. Then, the point clouds are divided into blocks with resolution $64 \times 64 \times 64$ and the 4000 largest blocks are selected.

Three different training sessions were carried out, using the five λ values described above. Figure 3 shows the results of the three training sessions, with zoomed areas of interest. The codec shows a high level of stability across the two additional training sessions, although in most cases, the D1 metric has a slight variation in the intermediate λ values. When analysing the training plots for *Citiusp*, the second and third training are highly similar when $\lambda = 3 \times 10^{-4}$ and 10^{-5}. For $\lambda = 5 \times 10^{-5}$, the D1 value for the first training session is above the achieved in the other two training sessions. The same can be observed when $\lambda = 10^{-4}$. This behaviour is found across the tested point clouds. Some small degrees of instability can also be found in the point clouds *Ramos* when $\lambda = 10^{-4}$ and in the point cloud *Guanyin* when $\lambda = 10^{-5}$.

Overall, this codec shows a higher degree of stability than PCGC. When encoding the *Guanyin*, the training sessions with $\alpha = 16$ reveal a small degree of instability. Other points where different training processes reach different Rate-Distortion relations are observed mostly to *Guanyin* and *Romanoillamp*.

However, these differences are very small and this codec always find working points that reveal a higher level of stability than PCGC.

D. Visual Examples

It is also important to visualize some examples to understand the variation of each training session in the decoded point clouds. Figure 4 shows a cropped area of the decoded *Guanyin* point clouds for each training session for both codecs. All training sessions of PCGC show some artifacts in the torso area. Small parts seem to be missing, always in different locations on the torso of the point cloud. In the face area, the second training session produced almost no artifacts, while the first and third training produced some holes. PCC_GEO_CNN reveal much more noticeable artifacts for the lower bit rate. All sessions produced a number of distortions in the face, especially training session two, where cracks can be seen in the nose area. In the torso area, some artefacts can be identified, like the ones found in PCGC, but in a larger scale.

Figure 5 also shows a cropped area of the decoded *Romanoillamp* point clouds for each training session for both codecs. PCGC creates an enormous amount of artifacts across the point cloud. The figure shows that it was impossible for the codec to find a good rate/distortion trade-off. All training sessions produced the same type of artifacts, PCC_GEO_CNN shows a similar behaviour to PCGC, albeit in a smaller scale. All training sessions created similar artifacts in the handle area of the point cloud. In the central area, some variation in the location of the artifacts can be observed.

IV. Discussion and conclusions

A stability analysis of the training process of two machine-learning based codecs, namely PCGC and PCC_GEO_CNN is reported. According to our analysis, both codecs show a high level of stability in the coding performance. PCC_GEO_CNN reveals the most stable across the training sessions. In the training of PCGC, the training dataset is randomly selected from the provided database, while in PCC_GEO_CNN, a static database is used. Given that, it was expected that PCC_GEO_CNN to be the most stable across the conducted training sessions.

When the desired rate-distortion points are reached for each codec, some instability is observed for both codecs. This effect is more visible for the PCGC. However, PCC_GEO_CNN has a stopping mechanism that prevents this effect to become visible. While PCGC has training session with a fixed number of 50 epochs, PCC_GEO_CNN never reaches that value. The number of training epochs is dynamically computed and the training session is stopped.

The observed instability near the limits of the training sessions might be due to some over-fitting mechanism, that does not allow to improve anymore the rate-distortion relation. This causes that different training sessions might have variable performance for the last training epochs of a training process, causing the observed instabilities for some content.

Moreover, it was also observed that different training sessions create different artifacts for the selected working points, although the PSNR D1 metric has similar values. The point to point metric is used for optimization, in both codecs leading to this result. However, it is likely to happen that different metrics can have different values. It is also important to understand if different training can lead to different perceptual quality. That requires to perform subjective evaluation, that has several problems that need to be considered. For instance, how to render the color in the point clouds, as that will have tremendous impact in the perceived quality and might mask any other quality analysis. The authors plan to consider this analysis in future studies.

Nevertheless, both codecs revealed a very reasonable stability on the performance for different training sessions, showing a high reliability. This is mostly observed for the PCC_GEO_CNN where the reached Rate-Distortion working points have only slight variations for different training processes. Nevertheless, PCGC also reveals a high level of stability, and can also be considered a reliable codec. It is expected that different training sessions will tend to create different artifacts, although the PSNR D1 metric is kept similar as it was used in the cost function for the codec optimisation.

References

[1] V. Zakharchenko, ""Algorithm description of mpeg-pcc-tmc2"," *ISO/IEC JTC1/SC29/WG11 MPEG2018/N17767*, Jul 2018.

[2] K. Mammou, P. A. Chou, D. Flynn, and M. Krivokuća, "G-PCC codec description v2," *ISO/IEC JTC1/SC29/WG11 N18189*, Jan 2019.

[3] S. Perry et al, "Quality evaluation of static point clouds encoded using mpeg codecs," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020.

[4] J. Prazeres, M. Pereira, and A. M. G. Pinheiro, "Quality analysis of point cloud coding solutions," in *2022 Electronic Imaging Symposium*, 2022.

[5] J. Wang, H. Zhu, H. Liu, and Z. Ma, "Learned point cloud geometry compression," *CoRR*, 2019.

[6] J. Wang, D. Ding, Z. Li, and Z. Ma, "Multiscale point cloud geometry compression," 2020.

[7] M. Quach, G. Valenzise, and F. Dufaux, "Learning convolutional transforms for lossy point cloud geometry compression," in *IEEE International Conference on Image Processing, ICIP*, 2019.

[8] ——, "Improved deep point cloud geometry compression," *CoRR*, vol. abs/2006.09043, 2020. [Online]. Available: https://arxiv.org/abs/2006.09043

[9] A. F. R. Guarda, N. M. M. Rodrigues, and F. Pereira, "Adaptive deep learning-based point cloud geometry coding," *IEEE Journal of Selected Topics in Signal Processing*, 2021.

[10] X. Sheng et al, "Deep-pcac: An end-to-end deep lossy compression framework for point cloud attributes," *IEEE Transactions on Multimedia*, 2021.

[11] L. Wiesmann et al, "Deep compression for dense point cloud maps," *IEEE Robotics and Automation Letters*, 2021.

[12] D. Tian et al, "Geometric distortion metrics for point cloud compression," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017.

[13] C. S. et al, "Inception-v4, inception-resnet and the impact of residual connections on learning," *CoRR*, 2016.

[14] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.

[15] A. X. Chang et al, ": An information-rich 3D model repository," 2015. [Online]. Available: https://arxiv.org/abs/1512.03012

[16] Z. Wu et al, "3D shapenets: A deep representation for volumetric shapes," 2014. [Online]. Available: https://arxiv.org/abs/1406.5670

5.6 Quality Evaluation of Machine Learning-based Point Cloud Coding Solutions

Quality Evaluation of Machine Learning-based Point Cloud Coding Solutions
J. Prazeres, Rafael Rodrigues, Manuela Pereira, and Antonio M.G. Pinheiro. 2022. Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis (APCCPA '22). Association for Computing Machinery, New York, NY, USA, 57–65.
DOI: https://doi.org/10.1145/3552457.3555730

Quality Evaluation of Machine Learning-based Point Cloud Coding Solutions

Joao Prazeres
joao.prazeres@ubi.pt
Universidade da Beira Interior and
Instituto de Telecomunicacoes
Covilhã, Portugal

Rafael Rodrigues
rafael.rodrigues@ubi.pt
Universidade da Beira Interior and
Instituto de Telecomunicacoes
Covilhã, Portugal

Manuela Pereira
mpereira@di.ubi.pt
Universidade da Beira Interior and
Instituto de Telecomunicacoes
Covilhã, Portugal

Antonio M. G. Pinheiro
pinheiro@ubi.pt
Universidade da Beira Interior and
Instituto de Telecomunicacoes
Covilhã, Portugal

ABSTRACT

In this paper, a quality evaluation of three point cloud coding solutions based on machine learning technology is presented, notably, ADLPCC, PCC_GEO_CNN, and PCGC, as well as LUT_SR, which uses multi-resolution Look-Up Tables. Moreover, the MPEG G-PCC was used as an anchor. A set of six point clouds, representing both landscapes and objects were coded using the five encoders at different bit rates, and a subjective test, where the distorted and reference point clouds were rotated in a video sequence side by side, is carried out to assess their performance. Furthermore, the performance of point cloud objective quality metrics that usually provide a good representation of the coded content is analyzed against the subjective evaluation results. The obtained results suggest that some of these metrics fail to provide a good representation of the perceived quality, and thus are not suitable to evaluate some distortions created by machine learning-based solutions. A comparison between the analyzed metrics and the type of represented scene or codec is also presented.

CCS CONCEPTS

• **HCI design and evaluation methods**; • **Visualization design and evaluation methods**;

KEYWORDS

Point Clouds, Machine Learning, Quality evaluation, Coding

ACM Reference Format:
Joao Prazeres, Rafael Rodrigues, Manuela Pereira, and Antonio M. G. Pinheiro. 2022. Quality Evaluation of Machine Learning-based Point Cloud Coding Solutions. In *Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis (APCCPA '22), October 14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3552457.3555730

APCCPA '22, October 14, 2022, Lisboa, Portugal.

ACM ISBN 978-1-4503-9491-8/22/10…$15.00
https://doi.org/10.1145/3552457.3555730

1 INTRODUCTION

In the modern world, the need for 3D data formats is increasing for multiple applications notably virtual, augmented and mixed reality, computer graphics, gaming, 3D printing, construction, manufacturing, robotics, automation, medical applications, cultural heritage, remote sensing and geographical information systems, and also consumer and retail. This applications might strongly benefit from reliable point cloud technology, increasing its effectiveness, and improving the quality and user experiences. The models of data representation, notably models for point clouds representation and associated quality will play an important role, as 3D content usually leads to huge amounts of information.

Point cloud technology has emerged as a popular method for data representation. It consists in a set of Cartesian coordinates (x, y, z), containing a list of attributes associated with each element, such as RGB components, reflectance, physical sensor information or normal vectors. Point clouds allow accurate representation of objects or scenes, from any viewing position or distance, thus making them a very powerful representation model, extremely useful in VR/AR scenarios, environment mapping for autonomous driving and urban and landscape mapping. An accurate point cloud of a city, building or artefact can contain several millions of points, each with one or more associated attributes. Considering this, point cloud compression solutions are needed, in order to efficiently compress and decompress this type of data. MPEG developed two powerful coding solutions, notably the Video-based Point Cloud Compression (V-PCC) [53] and the Geometry-based Point Cloud Compression (G-PCC) [30]. A quality study of these two codecs is provided in [35]. Google developed DRACO [14] for meshes and point cloud representation. Point clouds are coded without losses, although Draco has means to point cloud resolution control, and it can not compete with the MPEG codecs in terms of quality vs bit rate [37].

In recent years, machine learning-based coding solutions have emerged as effective models to encode point clouds [15–17, 39], after revealing to be high performing architectures for image coding.

Multiscale Point Cloud Geometry Compression was proposed in [51, 52] using a Minkowski Engine [10] for sparse convolutions.

In [40, 41], the Deep Point Cloud Geometry Compression was presented. Adaptive Deep Learning Point Cloud Compression was presented in [17] and LUT_SR was proposed in [12], as a post processing step for G-PCC. In [43] a point cloud lossy attribute auto encoder is proposed, directly encoding and decoding attributes with the help of geometry.

Up to our knowledge, the quality evaluation of these coding solutions is based on objective metrics. These metrics are even likely to be used in the cost functions used for the training of the coding solution. Moreover, the type of errors caused by the machine learning solutions tends to be quite different of the caused by the common codecs, as typically machine learning based codecs tend to create holes in the point clouds surface. Because of that, a suitable subjective evaluation that compares different machine learning based coding solutions of point clouds is of upmost importance, as it will provide a reliable comparison between different solutions, and will provide a reliable benchmarking for different point cloud quality metrics.

In this work, a quality evaluation of three machine learning-based coding solutions is presented, notably, Adaptive Deep Learning Point Cloud Coding (ADLPCC) [17], Multiscale Point Cloud Geometry Compression (PCGC) [51], Deep Point Cloud Geometry Compression (PCC_GEO_CNN) [40]. Another recent solution, which uses Look-Up Tables built on geometric similarities across scales to resolve low resolution point clouds (LUT_SR) [12] was evaluated. Finally, the MPEG G-PCC [30] was used as benchmarking anchor.

Unlike most of previous studies that only study the quality of small objects, point cloud representatives of landscapes were also included in this paper. Furthermore, the current evaluation maps the color attributes in the coded point cloud. This is needed because the tested codecs only encode geometry. However, the texture created by color attributes plays an important role in the perceived quality.

Point cloud quality evaluation methodologies were several times studied. In [6] geometry only point clouds are considered and quality models are established. Compression artifacts using prior encoding schemes are evaluated in [11, 24, 34]. Current efforts account for a wider range of high-performing codecs, such as the ones reported in [7, 35, 44]. In [35] a subjective quality evaluation using MPEG Point Cloud codecs is presented. This work also considers a set of point cloud metrics, concluding that point to point and point to plane metrics [47] are the best performing ones, and provide a good representation of the subjective evaluation. In [6], a subjective evaluation was conducted in which point clouds were coded with octree based compression, and displayed with Screened Poisson surface reconstruction. In [11], a set of point clouds were coded using octree-pruning and a projection based method, with three levels of degradation. The models were displayed using point sizes large enough to ensure a visualisation of watertight surfaces. In [36], crowd sourcing was employed in subjective evaluation. The participants were given the option of downloading the subjective experiment content, or to access an online server and conduct the evaluation on a web browser. The two types of subjective evaluation reveal a very high level of statistical similarity. In [3], a subjective evaluation using AR is proposed, whereas in [45], a VR environment is used to evaluate point clouds. In [5], a point cloud toolbox was created, in order to aid subjective testing in such environment. In [31], users represent by 3D point clouds were able to interact with a virtual room. In [54], different resolutions and noise types were considered in a VR subjective evaluation. In [22], an evaluation on point cloud denoising algorithms is proposed. In [38], a subjective quality evaluation is conducted in a 3D environment. The results were compared to a previous study using 2D displays, and revealed that there were no statistical differences between the results. Additionally, objective quality evaluation aims to propose algorithms which can accurately predict the visual quality of content representations. Having access to algorithms that can accurately predict the quality of coded contents it's incredibly valuable as they will allow reliable quality estimation without the need of subjective quality evaluations or to easily setup codecs for improved quality of experience. Additionally, the benchmarking of these solutions is facilitated by using the best objective quality metrics, replacing the need to conduct subjective quality evaluations. In [29], a database for point cloud quality assessment is proposed together with a subjective evaluation.

Objective quality metrics can be divided in image-based and model based [27]. The first exploits high-performing solutions, applied afterwards on the selected representative views of the model, while the second one relies on geometric error, curvature or statistics measures [26], among others. The most common model-based approaches mainly assess geometry and rely on Euclidean distances, or projected errors along normal vectors [48]. In [1], an algorithm based on local surface approximations was proposed and in [2], an algorithm based on geometry normals, curvature statistics and color was introduced. In [32], color errors based on MSE and PSNR are applied on either RGB or YCbCr color space. In [50], histograms, representing color statistics are used to predict texture distortion of point cloud contents. A broad study of objective quality metrics was conducted in [28]

In the following section, a short description of the tested codecs is presented. In section 3, data preparation is described. The subjective and objective quality evaluation are described in section 4 and 5, respectively, along with a discussion of the obtained results.

2 CODEC DESCRIPTION

In this section, the codecs considered for this study are shortly described.

2.1 Adaptive Deep Learning Point Cloud Coding

The Adaptive Deep Learning Point Cloud Coding (ADLPCC) [17] initialially creates a partition of the point cloud into regular-sized 3D blocks, which are encoded and reconstructed separately by several models. The common architecture of the encoding models consists of an autoencoder (AE) and a variational autoencoder (VAE), each with three convolutional layers for encoding and three for reconstruction, with sigmoid and ReLU activations, respectively. The resulting latent representations of the AE and VAE are both entropy coded.

Each coding model is evaluated on a given block using objective PC distortion metrics and bit rate measurements. The point cloud decoding uses the decoding counterparts of the selected model for a given block, before being reconverted into 3D coordinates and

merging. ADLPCC uses a loss function that takes into account both the distortion (D) and the block coding rate (R), according to:

$$J = D + \lambda R \tag{1}$$

Varying the hyperparameter λ allows to obtain different rate-distortion trade-offs. The block distortion in measured using a focal binary cross-entropy function (Eq. 2), which incorporates two training hyperparameters α and γ to adapt to block sparsity characteristics.

$$FL(v,u) = \begin{cases} -\alpha(1-v)\gamma\log(v) & u=1 \\ -(1-\alpha)v^{\gamma}\log(1-v) & u=0 \end{cases} \tag{2}$$

In Eq. 2, u and v refer to the original voxel occupancy binary value and the reconstructed voxel probability, respectively.

2.2 Multiscale Point Cloud Geometry Compression

The Multiscale Point Cloud Geometry Compression (PCGC) model [51] performs block-wise multiresolution encoding. Each encoding module consists of two simple convolutional layers for downscaling, followed by three residual feature extraction blocks, each containing three instances of the Inception Residual Network [46]. The downsampling is performed three times, resulting in different representations (X', X", and Y). At the bottleneck, the geometry coordinates (C_Y) and feature attributes (F_Y) of the latter are encoded, using the octree codec [30] and entropy coding, respectively. The decoding branch mirrors the encoding part, with the decoded C and F components as input, and includes a hierarchical classification at each scale.

2.3 Deep Point Cloud Geometry Compression

Deep Point Cloud Geometry Compression (PCC_GEO_CNN) [40] aims at reducing the blocking effect introduced by other deep learning-based codecs, by taking the original point cloud as input. PCC_GEO_CNN learns an encoding function from three sequential convolutional layers. While the first two layers use ReLU activation, the latent representation from the third layer ($y = f_a(x)$) is quantized $\hat{y} = Q(y)$, using element-wise integer rounding. $\hat{y}$ is compressed through range coding and the Deflate algorithm, which is a combination of LZ77 and Huffman coding [20] with shape information on the reference cloud x and latent representation y added before compression.

The decoding function (f_s) mirrors f_a, with three transposed convolutional layers, all using a ReLU activation function. The last layer has only 1 filter and its output is converted into the distorted point cloud ($\tilde{x}$) using element-wise minimum, maximum and rounding functions. The global loss function is similar to Eq. 1, but with the λ parameter associated to the distortion. The distortion is, in turn, computed using the focal loss defined in Eq. 3, to compensate the larger number of empty voxels:

$$FL(p_z^t) = -\alpha_z(1-p_z^t)^{\gamma}\log(p_z^t) \tag{3}$$

In this loss function, if the voxel z is occupied, p_z^t and α_z are defined as p_z (reconstructed voxel probability) and α, respectively. Otherwise, they are defined as $1-p_z$ and $1-\alpha$. The parameters α and γ are hyperparameters of the model training.

2.4 Look-Up Tables

This solution creates a hierarchical tree-like dictionary, named Look-Up Table (LUT), which maps the occupancy relationships between downsampled geometries (V_d) and its originating counterparts (V) [12]. The method performs a second downsampling step taking V_d as input geometry, and a fractional scale factor s, resulting in the parent geometry V_{d^2}. Then, the child occupancy for each parent voxel, $\sigma(v_{d^2}(k))$, may be defined and stored in the LUT. The neighborhood configuration $\varphi_M(v_{d^2}(k))$, where M defined a M^3 cude around $v_{d^2}(k)$ is also stored in the LUT. Using a fractional scale avoids dealing with irregular grids. However, the resulting downsampled geometries may then have different configurations/classifications, because each coordinate (x,y,z) in V_d could be uniparous or multiparous, according to the number of corresponding children. The LUT will have m entries, one for each possible geometry configuration, that estimates the most likely child occupancy for the neighborhood $\varphi_M(m)$:

$$\overline{\sigma}(m) = E\{\sigma(v_{d^2}(k)) \mid \varphi_M(m)\} \tag{4}$$

The upsampling stage to obtain a resolved point cloud (V_{sr}) first applies the nearest-neighbor interpolation to find all the possible child nodes of the input point cloud V_d. The resulting geometry is carved to obtain V_{sr} following the respective LUT entries, i.e., for the corresponding φ_M, to know what points to remove. Color interpolation for the resolved point cloud takes the weighted average of the adjacent neighbors, whose weights are scale-dependent.

2.5 Geometry Point Cloud Compression

The Geometry Point Cloud Compression (G-PCC) codec is one of the normalized MPEG point cloud codecs [30], and is used as an anchor in this study. Geometry encoding in G-PCC may follow one of two methods of point cloud compression, notably the octree based method, and the trisoup method, which is based on surface reconstruction using triangular primitives, after octree decomposition. For this study, only the octree method was considered as it usually presents a more stable behaviour, and the performance of both methods is not significantly different, as shown in [35].

The G-PCC loss is controlled by the positionQuantizationScale (pQs) parameter, that controls the number of divisions of the octree, from the root to each leaf node, leading to a regular downsampling of the input point clouds. Geometry is first encoded using the octree method, and then decoded to define the shape over which the color will lie. The color attributes are assigned to output points through a re-colouring step, which uses the color values of the original model. G-PCC uses one of two different approaches to encode the color information, namely RAHT [13], based on the 3D Haar transform, and Prediction-plus-Lifting [4], based on prediction of a color value from its neighbours. For this study, only the Prediction-plus-Lifting method was considered, as it was shown in [7] that users tend to prefer the Lifting codec over RAHT. It uses a QP parameter that controls the losses on the texture information. The point clouds were encoded with the parameters shown in table 1.

Table 1: G-PCC encoding parameters.

Rate	R01	R02	R03	R04	R05
QP	46	40	34	28	22
pQS	0.25	0.5	0.75	0.875	0.9375

3 DATA SELECTION AND PREPARATION

A set of six point clouds available at JPEG Pleno database[1] was selected, depicting three objects and three landscapes. The objects are the *Romanoillamp* from Univ. of Sao Paulo Database[2], the *Guanyin* from the EPFL dataset, and frame 1300 of the *Longdress* dynamic point cloud. The point clouds *Citiusp*, *IpanemaCut* and *Ramos*, from Univ. of Sao Paulo Database were selected. Figure 1 represents the chosen point clouds.

Prior to the test, the point clouds were voxelized [1, 5] by quantizing the coordinates of the models and blending the colors of points in the same voxel. A voxel depth of 10 was empirically chosen, to ensure that each point could be represented by one pixel of the 4K display used for evaluation, and used for all the point clouds in the subjective test dataset.

Regarding video preparation, several point cloud views were captured using PCLVisualizer [42], each representing a 1° rotation. A complete rotation about the vertical axis was depicted, and the full sequence was rendered at 30 fps, thus resulting in 12 second videos. These video sequences were created with FFMPEG [3], using the H.264 codec[18].

The Constant Rate Factor (CRF) and q parameters were set to 0, so that no compression was applied. Furthermore, to prevent any RGB to YUV colorspace conversion the `libx264rgb` option was also used. In some cases, the point size was changed to provide an improved visual representation (Table 2), maintaining surfaces continuous as much as possible. If transparency appears in the point cloud, the subjects would likely see the opposite part of the point cloud, leading to a very bad quality perception [6, 11].

The machine learning-based codecs used in this study only encode geometry information. Thus, to prepare the distorted contents for the subjective evaluation, the color information of the reference point clouds was mapped onto the corresponding distorted point cloud using *Meshlab* [4]. Texture is very important in the definition of the perceived quality. It was also observed for the definition of perceived quality, to balance the quality of texture with the quality of the geometry. Because of that, the point clouds were encoded with G-PCC using the `lossless-geometry-lossy-atts` mode. For each rate in the subjective test (R01-R05), the QP value is set to the corresponding value in table 1, with a fixed pQs value of 1.

4 SUBJECTIVE QUALITY EVALUATION

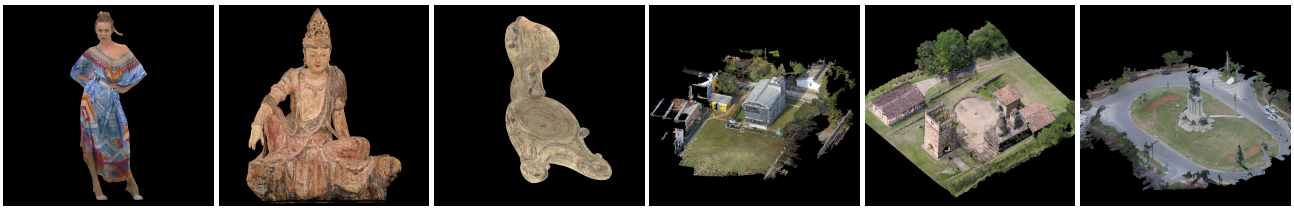
A total of 20 participants were involved in the subjective study, with ages between 18 and 58 (24.7±8.3), from which 15 were male and 5 were female. The subjective test setup used a 31.1 inch Eizo ColorEdge CG318-4K, with a full resolution of 4096x2160, and followed the specifications in [9]. During the test, each participant was shown a randomized sequence of videos containing both the quality reference and a distorted version, side by side. To avoid biases, half of the subjects were shown videos with the reference on the right and the distorted content on the left, and the other half vice-versa. A Double Stimulus Impairment Scale was used, with the subjects being prompted to evaluate the quality of the distorted point cloud, in comparison to the provided reference, according to a five-level rating scale (1 - very annoying, 2 - slightly annoying, 3 - annoying, 4 - perceptible, but not annoying, 5 - imperceptible). After the subjective test, the Mean Opinion Score (MOS) for every content was computed, by taking the average of their obtained scores. The coding bit rates, measured in bits per point (bpp), were computed as $bpp = n\_bits_d / n\_points_o$, where n_bits_d is the number of bits of a distorted content, and n_points_o is the number of points of the original point cloud.

Five quality levels were considered for each codec-content pair (Table 2), giving a total of 150 videos. Moreover, hidden reference-reference pairs for every content were also included in the test sequence, raising the final total to 156 videos. Distorted versions of the same content were never shown back to back. Before the proper subjective test, subjects were shown a training sequence with eight videos, to become familiarized with the distortion artifacts typically created by the codecs. These included four degradation levels for two different point clouds, namely *Airplane* (after conversion from a mesh) from the PointNet Database, and *Villalobospark*, both from the University of São Paulo dataset.

Figure 2 shows the MOS obtained in the subjective tests plotted against the respective coding bit rates (bpp), for each point cloud in the test dataset. Although the opinion scores do not have a gaussian distribution, the 95% Confidence Interval (CI) was computed, assuming a Student's t-distribution. The horizontal green line at the top of each plot refers to the MOS for the hidden references, whereas the green bar around it represents its 95% CI. The vertical black bar at the right side of each plot, represents the lossless encoding with G-PCC. This was computed to assure that the tested bit rates were not greater than the lossless bit rate of G-PCC. It should be noted here that G-PCC and LUT_SR were tested for similar bit rates, allowing a more direct comparison. For learning-based codecs, i.e., ADLPCC, PCC_GEO_CNN, and PCGC, the resulting bit rates are highly dependent on their training, thus the same was not possible. Nevertheless, these are directly comparable within their selected range of bit rates, and are simultaneously comparable to the higher bit rates of G-PCC and LUT_SR.

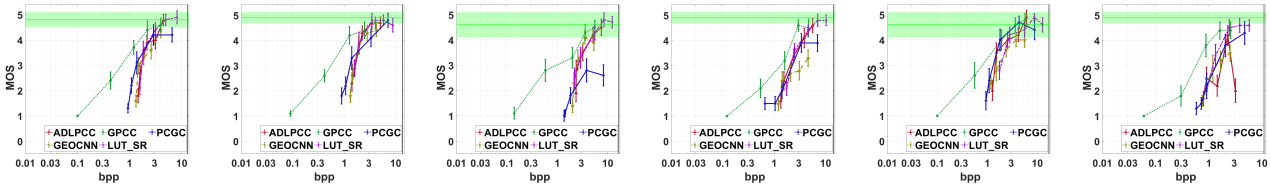
None of the learning-based codecs, was able to reach the performance of the anchor G-PCC. Globally, all three codecs showed a very similar performance, regardless of the content, with only a few exceptions. PCGC only reaches a MOS similar to the reference for the *IpanemaCut* (R04 and R05) and the *Guanyin* (R05) point clouds. However, it seems to have a slighty better performance for low bit rates than ADLPCC and PCC_GEO_CNN for some content, namely *Longdress*, *Guanyin*, and *IpanemaCut*. The *Romanoillamp* point cloud is an outlier to the general behavior. PCGC performed quite poorly for the *Romanoillamp* point cloud, as the MOS never reached 3 for any bit rate, which may be related to a lack of suitable data in the training set. Excerpts of the resulting coded point cloud may be seen in Figure 6. In the case of ADLPCC, the *Ramos* point cloud even reveals a strange behavior, where the higher bit rate

[1] http://plenodb.jpeg.org/pc/8ilabs
[2] http://uspaulopc.di.ubi.pt
[3] https://ffmpeg.org/
[4] https://www.meshlab.net/



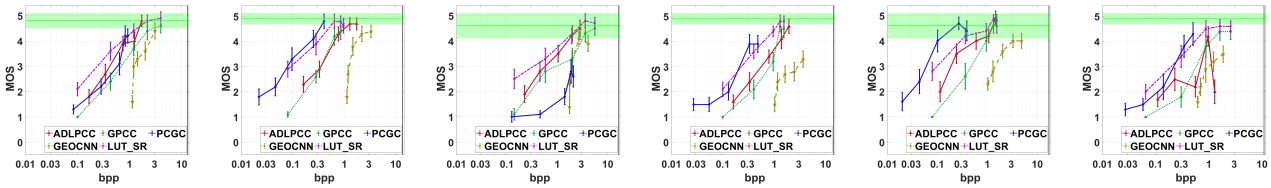
(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Figure 1: Point Cloud test set.



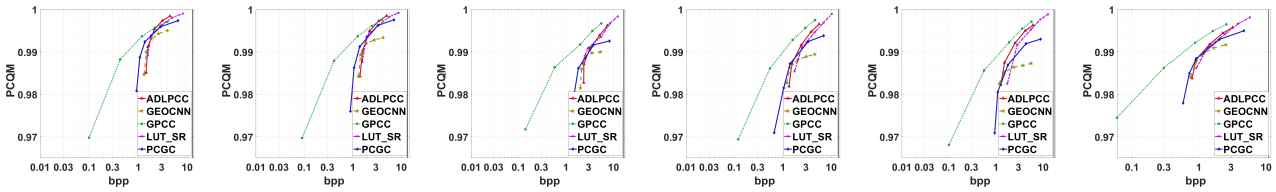
(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Figure 2: MOS vs bpp with 95% confidence interval considering both texture (encoded with G-PCC) and geometry.



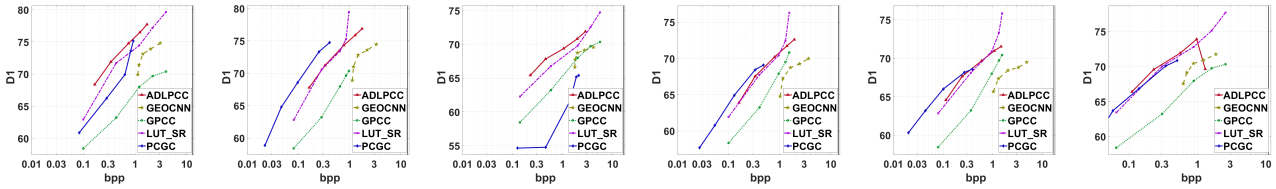
(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Figure 3: MOS vs bpp with 95% confidence interval considering the geometry information only.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Figure 4: PCQM vs bpp.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *IpanemaCut* (f) *Ramos*

Figure 5: D1 vs bpp.

Table 2: Point size of each point cloud for visualization in the subjective test.

	ADLPCC					PCC_GEO_CNN					LUT_SR					G-PCC					PCGC				
Content	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05
Longdress	7	4	3			8	6	5	3		3					8	4	3			12	10	9	3	
Guanyin	7	4	3			8	6	5	3		3					8	4	3			12	10	9	3	
Romanoillamp	15	6	5			10	8	7	6		5					10	6	5			15	14	13	8	7
Citiusp	3					3					3					12	6	5	3		3				
IpanemaCut	3					3					3					12	6	5	3		3				
Ramos	3					3					3					12	6	4	3		3				

results in very low performance. The reason can be observed in Figure 7, where a significant part of the point cloud disappeared. This might also be caused by a lack of suitable training data.

Fig. 3 shows the plots of the MOS obtained as a function of the bit rate of the geometry only, without the texture influence. This plots are important because the tested machine learning-based codecs only encode the geometry.

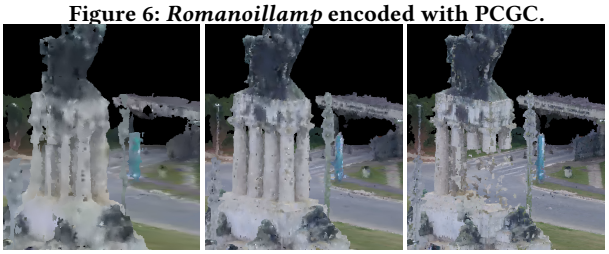
Apart the PCC_GEO_CNN codec, in general all the encoders lead to a better MOS than the anchor G-PCC, when considering the geometry only. Their performance only becomes worst when the texture is added. In particular, the codec LUT_SR successfully improved the performance of G-PCC on geometry, but that improvement is lost when the texture is added.

Another unusual behaviour was observed for the higher bit rate of ADLPCC (R05), which yielded a very low perceived quality for the *Ramos* point cloud, following a MOS above 4 with R04. This might also be caused by a lack of appropriate training data. An excerpt of the resulting coded point may be seen in Figure 7. ADLPCC also revealed some blocking artifacts, due to its initial point cloud partition, that may have further influenced its scores.

As shown in Table 2, for the content representing landscapes, only G-PCC required as increase of the point size for the lower bit rates. All other codecs kept a good integrity of the surfaces in visualization using the standard point size, without a major decay in the experienced quality. In the case of content representing objects, most codecs required adjustment, especially in lower bit rates, with the exception of LUT_SR. Fig. 6 and 8 represent an example for PCGC and PCC_GEO_CNN.

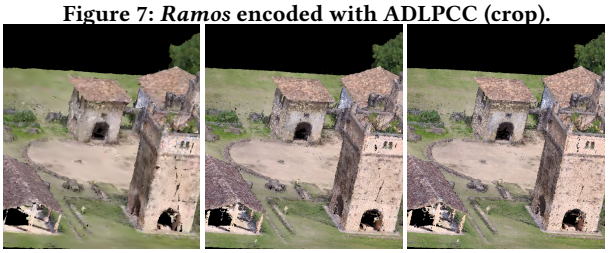
(a) *R01* **(b) *R03*** **(c) R05**

Figure 6: *Romanoillamp* encoded with PCGC.



(a) R01 **(b) R03** **(c) *R05***

Figure 7: *Ramos* encoded with ADLPCC (crop).



(a) R01 **(b) R03** **(c) *R05***

Figure 8: *IpanemaCut* encoded with PCC_GEO_CNN (crop).

5 OBJECTIVE QUALITY EVALUATION

Subjective quality assessment provides a ground truth for the validation of objective quality metrics, in the presence of the distortions produced by these codecs. In this paper, the performance of a selected set of objective quality metrics are described, notably the D1 and D2 metrics [48], Point cloud Structural Similarity metric (PSSIM) using color attributes and the covariance estimator [5], Point 2 Distribution metric [23, 25], Point Cloud Metric - Reduced Reference (PCM-RR) [49], and Point Cloud Quality Metric (PCQM) [32], by correlating their predicted MOS with the subjective MOS. The metrics selection is based on past experiences.

The MOS predictions of a given metric were computed after logistic regression on the objective scores, as is commonly done when benchmarking objective metrics [19, 33]. Then, the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Root-Mean Squared Error (RMSE), and the Outlier Ratio (OR) were computed to measure the correlation, as specified in [33]. The Cloud Compare Quadric Fitting with a radius of 5 [8] was used to compute the normals [21] in the cases it was needed.

Table 3 shows the correlation outcomes of each metric, either for the entire dataset (*Global*) or by type of content (*Landscapes* or *Objects*). PCQM achieved the best results by a large margin, PCC = 0.911 and SROCC = 0.912 when considering the entire test set. It was the only metric reaching correlation coefficients above 0.9, even when considering types of content separately (PCC/SROCC of 0.932/0.928, and 0.910/0.906, for landscapes and objects, respectively). PSSIM and Point 2 Distribution provided the next best

Table 3: Correlation of the objective metrics with the subjective MOS. Results under *Global* refer to the MOS for all codecs and content types, while results under *Landscapes* and *Objects* refer to each content type separately.

	Global				Landscapes				Objects			
Metric	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
PCQM	**0.911**	**0.912**	**0.127**	**0.547**	**0.932**	**0.928**	**0.114**	**0.140**	**0.910**	**0.906**	**0.124**	**0.570**
D1 MSE PSNR	0.820	0.797	0.175	0.780	0.843	0.841	0.165	0.600	0.822	0.783	0.171	0.660
D2 MSE PSNR	0.824	0.801	0.173	0.787	0.835	0.829	0.169	0.560	0.817	0.784	0.173	0.750
Point 2 Distribution	0.853	0.847	0.160	0.673	0.864	0.851	0.155	0.520	0.850	0.849	0.159	0.570
PSSIM	0.855	0.858	0.159	0.687	0.836	0.831	0.169	0.560	0.862	0.867	0.152	0.620
PCM-RR	0.831	0.834	0.171	0.855	0.887	0.885	0.142	0.440	0.814	0.826	0.175	0.750

Table 4: Correlation of the objective metrics with the subjective MOS for each codec.

	ADLPCC				PCC_GEO_CNN				LUT_SR				G-PCC				PCGC			
Metric	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
PCQM	0.854	0.843	0.146	0.300	0.798	0.774	0.141	0.767	**0.950**	**0.895**	**0.075**	**0.200**	**0.984**	**0.956**	**0.063**	**0.100**	0.889	0.874	0.147	0.233
D1 MSE PSNR	0.788	0.763	0.173	0.533	0.683	0.644	0.171	0.633	0.939	0.850	0.083	0.267	**0.984**	**0.938**	**0.064**	**0.100**	0.915	0.899	0.128	0.300
D2 MSE PSNR	0.826	0.814	0.159	0.367	0.819	0.820	0.134	0.433	0.946	0.850	0.079	0.017	0.980	0.941	0.071	0.100	0.834	0.817	0.716	0.400
Point 2 Distribution	**0.894**	**0.903**	**0.126**	**0.233**	**0.925**	**0.927**	**0.088**	**0.267**	**0.950**	**0.903**	**0.075**	**0.167**	0.758	0.660	0.240	0.400	**0.933**	**0.925**	**0.114**	**0.233**
PSSIM	0.869	0.875	0.139	0.300	0.906	0.908	0.099	0.300	0.915	0.876	0.097	0.167	0.935	0.929	0.127	0.200	0.907	0.896	0.135	0.300
PCM-RR	0.753	0.784	0.185	0.533	0.719	0.707	0.163	0.533	0.807	0.803	0.142	0.433	0.967	0.938	0.091	0.133	**0.934**	**0.917**	**0.113**	**0.300**

performances, and quite similar between them, considering the entire test set, with PCC/SROCC of 0.855/0.858, and 0.853/0.847, respectively.

When considering the correlations using only point clouds of landscapes, PCM-RR shows an improvement of its performance, reaching a PCC of 0.887 and SROCC of 0.885, which makes it the second best metric for landscape point clouds. Indeed, the same occurred with D1 and D2, albeit on a smaller scale, as these performed better with point clouds of landscapes. In the case of objects, D1 and D2 (PCC of 0.822 and 0.817) yielded similar PCC values to PCM-RR (0.814). However, their rank correlation coefficient even dropped below 0.8 (SROCC of 0.783 and 0.784).

Table 4 shows the performance of the objective metrics for each codec, while Fig. 9 presents the corresponding normalized objective metric vs. normalized MOS plots. For machine learning-based codecs i.e., ADLPCC, PCC_GEO_CNN, and PCGC, Point 2 Distribution is consistently the best performing metric, with PCC/SROCC of 0.894/0.903, 0.925/0.927, and 0.933/0.925, respectively. Interestingly, PCM-RR, which had a poor performance for ADLPCC and PCC_GEO_CNN, similar to Point 2 Distribution for PCGC, obtaining PCC = 0.934 and SROCC = 0.917. PSSIM held the second best performance for ADLPCC and PCC_GEO_CNN, and also achieved a good performance for PCGC. PCQM and D1 also had reasonable performances with PCGC.

Point 2 Distribution was again the best metric for LUT_SR (PCC/SROCC = 0.950/0.903), close to PCQM (PCC/SROCC = 0.950/0.895). However, it was the worst performing metric for the G-PCC anchor codec, with a PCC below 0.8. As was expected, PCQM, D1 and D2 performed well with G-PCC, always reaching coefficients above 0.9.

Figure 4 and 5 show the PCQM [32] and D1 metric [47] performances, respectively, plotted against the coding bit rates, for each point cloud on the dataset. Figure 4 show the geometry plus texture bitrates, as this metric considers both geometry and color.

Figure 4 shows the GPCC codec obtains the best results. The ADLPCC seems to perform the best different between deep learning coding solutions. However, PCGC seems to perform better in the low bit rates. Moreover, PCQM does not reveal the bad behavior of the ADLPCC for *Ramos* point cloud.

Figure 5 does not use texture information. As such, the plots show the geometry bitrates only. This metric reveals different performance for the different codecs. Although this metric reveals to be quite reliable for the analysis of some individual codecs, it can not be used to compare different coding solutions.

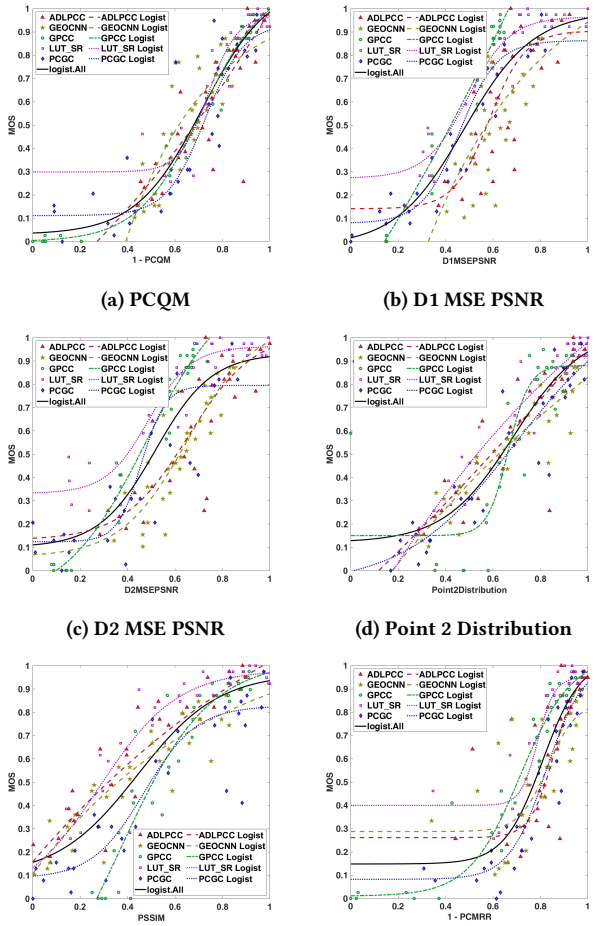
6 CONCLUSIONS

A study on the performance of static point clouds machine learning-based codecs was reported, notably, ADLPCC, PCC_GEO_CNN, and PCGC. This new generation of point cloud codecs is seen as a possible way to provide efficient compression beyond state-of-the-art codecs. Moreover, a very recent coding framework, based on self-similarities across different resolution's, i.e., LUT_SR was also tested. The G-PCC codec was also used as anchor.

From the presented subjective study, it can be concluded that most of the machine learning-based codecs have very similar performances. The PCGC seems to have some advantage in the lower bit rates, but it performs worst in the higher bit rates. Furthermore, they are more efficient than G-PCC in representing the geometry, with the exception of the PCC_GEO_CNN.

As texture is very important in any subjective evaluation, we made the option of encoding texture with G-PCC, keeping the same attribute quality parameter, as it is typically done with G-PCC. When the texture is incorporated, the anchor G-PCC has a much better performance than the other codecs. This indicates that coding solutions incorporating texture are also required for the future.

The tested machine learning-based codecs tend to create surface regions without points, which seem to produce a very strong



(a) PCQM (b) D1 MSE PSNR

(c) D2 MSE PSNR (d) Point 2 Distribution

(e) PSSIM (f) PCM-RR

Figure 9: Objective metric vs. MOS plots, with logistic regression curves (global and for each codec).

bad quality perception. This was specially noted in the *Romanoil-lamp* point cloud, encoded with PCGC and the *Ramos* point cloud, encoded with ADLPCC.

Regarding the objective quality evaluation, it was observed that the tested metrics have some problems in predicting the perceived quality of point clouds encoded with machine learning-based codecs.

ACKNOWLEDGMENTS

This research was funded by the Portuguese FCT-Fundação para a Ciência e Tecnologia under the project UIDB/50008/2020, PLive X-0017-LX-20, and by operation Centro-01-0145-FEDER-000019 - C4 - Centro de Competencias em Cloud Computing.

REFERENCES

[1] E. Alexiou and T. Ebrahimi. 2018. Point Cloud Quality Assessment Metric Based on Angular Similarity. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*.

[2] Evangelos Alexiou and Touradj Ebrahimi. 2020. Towards a Point Cloud Structural Similarity Metric. In *2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*.

[3] Evangelos Alexiou, Evgeniy Upenik, and Touradj Ebrahimi. 2017. Towards subjective quality assessment of point cloud imaging in augmented reality. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*.

[4] Evangelos Alexiou, Irene Viola, and Pablo Cesar. 2021. PointPCA: Point Cloud Objective Quality Assessment Using PCA-Based Descriptors. arXiv:2111.12663 [cs.MM]

[5] Evangelos Alexiou, Nanyang Yang, and Touradj Ebrahimi. 2020. PointXR: A Toolbox for Visualization and Subjective Evaluation of Point Clouds in Virtual Reality. *2020 Twelfth International Conference On Quality Of Multimedia Experience (Qomex)*.

[6] Evangelos Alexiou et al. 2018. Point Cloud Subjective Evaluation Methodology based on 2D Rendering. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*.

[7] E. Alexiou et al. 2019. A comprehensive study of the rate-distortion performance in MPEG point cloud compression. (2019).

[8] Pengbo Bo, Ruotian Ling, and Wenping Wang. 2012. A revisit to fitting parametric surfaces to point clouds. *Computers & Graphics* (2012). Shape Modeling International (SMI) Conference 2012.

[9] ITU-R BT.500-13. 2012. Methodology for the subjective assessment of the quality of television pictures,.

[10] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 2019. 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[11] L. A. da Silva Cruz et al. 2019. Point cloud quality evaluation: Towards a definition for test conditions. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*.

[12] Ricardo de Queiroz, Diogo Garcia, and Tomas Borges. 2021. Fractional Super-Resolution of Voxelized Point Clouds.

[13] Ricardo L. de Queiroz and Philip A. Chou. 2016. Compression of 3D Point Clouds Using a Region-Adaptive Hierarchical Transform. *IEEE Transactions on Image Processing* (2016).

[14] Google. 2020. Draco PCC Software. Retrieved February 28, 2021 from https://github.com/google/draco

[15] André F. R. Guarda, Nuno M. M. Rodrigues, and Fernando Pereira. 2019. Point Cloud Coding: Adopting a Deep Learning-based Approach. In *2019 Picture Coding Symposium (PCS)*.

[16] André F. R. Guarda, Nuno M. M. Rodrigues, and Fernando Pereira. 2020. Deep Learning-based Point Cloud Geometry Coding with Resolution Scalability. In *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*.

[17] André F. R. Guarda, Nuno M. M. Rodrigues, and Fernando Pereira. 2021. Adaptive Deep Learning-Based Point Cloud Geometry Coding. *IEEE Journal of Selected Topics in Signal Processing* (2021).

[18] ITU H.264. 2021. Recommendation H.264.

[19] Philippe Hanhart et al. 2015. Benchmarking of objective quality metrics for HDR image quality assessment. *EURASIP Journal on Image and Video Processing* (2015).

[20] David A. Huffman. 1952. A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the IRE* 40, 9 (1952), 1098–1101. https://doi.org/10.1109/JRPROC.1952.273898

[21] ISO/IEC JTC1/SC29/WG1. 2020. JPEG Pleno PC Exploration Study 4 Results. WG1M89044, 89th Meeting.

[22] Alireza Javaheri, Catarina Brites, Fernando Pereira, and João Ascenso. 2017. Subjective and objective quality evaluation of 3D point cloud denoising algorithms. In *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*.

[23] Alireza Javaheri, Catarina Brites, Fernando Pereira, and João Ascenso. 2021. A Point-to-Distribution Joint Geometry and Color Metric for Point Cloud Quality Assessment.

[24] A. Javaheri et al. 2017. Subjective and objective quality evaluation of compressed point clouds. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*.

[25] Alireza Javaheri et all. 2020. Mahalanobis Based Point to Distribution Metric for Point Cloud Geometry Quality Evaluation. *IEEE Signal Processing Letters* (2020).

[26] Guillaume Lavoué. 2011. A Multiscale Metric for 3D Mesh Visual Quality Assessment. *Computer Graphics Forum* (2011).

[27] Guillaume Lavoué, Mohamed Chaker Larabi, and Libor Váša. 2016. On the Efficiency of Image Metrics for Evaluating the Visual Quality of 3D Models. *IEEE Transactions on Visualization and Computer Graphics* (2016).

[28] Davi Lazzarotto, Evangelos Alexiou, and Touradj Ebrahimi. 2021. Benchmarking of objective quality metrics for point cloud compression. In *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*. 1–6. https://doi.org/10.1109/MMSP53017.2021.9733538

[29] Qi Liu, Honglei Su, Zhengfang Duanmu, Wentao Liu, and Zhou Wang. 2022. Perceptual Quality Assessment of Colored 3D Point Clouds. *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–1. https://doi.org/10.1109/TVCG.2022.3167151

[30] K. Mammou, P. A. Chou, D. Flynn, and M. Krivokuća. 2019. G-PCC codec description v2. *ISO/IEC JTC1/SC29/WG11 N18189* (Jan 2019).

[31] Rufael Mekuria, Kees Blom, and Pablo Cesar. 2017. Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video. *IEEE Transactions on Circuits and Systems for Video Technology* (2017).

[32] Gabriel Meynet et al. 2020. PCQM: A Full-Reference Quality Metric for Colored 3D Point Clouds. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*.

[33] ITU-T P.1401. 2012. International Telecommunication Union,. In *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*.

[34] S. Perry et al. 2019. Study of Subjective and Objective Quality Evaluation of 3D Point Cloud Data by the JPEG Committee. *Electronic Imaging* 2019, 10 (2019).

[35] Stuart Perry et al. 2020. Quality Evaluation Of Static Point Clouds Encoded Using MPEG Codecs. In *2020 IEEE International Conference on Image Processing (ICIP)*.

[36] Stuart Perry et al. 2021. Comparison of Remote Subjective Assessment Strategies in the Context of the JPEG Pleno Point Cloud Activity. In *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*.

[37] João Prazeres, Manuela Pereira, and António M. G Pinheiro. 2022. Quality analysis of point cloud coding solutions. In *2022 Electronic Imaging Symposium*.

[38] João Prazeres, Manuela Pereira, and António M. G Pinheiro. 2022. Subjective Quality Evaluation Of Point Clouds With 3D Stereoscopic Visualization. In *IEEE International Conference on Image Processing (ICIP)*.

[39] Maurice Quach, Giuseppe Valenzise, and Frédéric Dufaux. 2019. Learning Convolutional Transforms for Lossy Point Cloud Geometry Compression. *CoRR* (2019).

[40] Maurice Quach, Giuseppe Valenzise, and Frédéric Dufaux. 2019. Learning Convolutional Transforms for Lossy Point Cloud Geometry Compression. In *2019 IEEE International Conference on Image Processing, ICIP 2019, Taipei, Taiwan, September 22-25, 2019*.

[41] Maurice Quach, Giuseppe Valenzise, and Frederic Dufaux. 2020. Improved Deep Point Cloud Geometry Compression. arXiv:2006.09043 [cs.CV]

[42] Radu Bogdan Rusu and Steve Cousins. 2011. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Shanghai, China.

[43] Xihua Sheng et al. 2021. Deep-PCAC: An End-to-End Deep Lossy Compression Framework for Point Cloud Attributes. *IEEE Transactions on Multimedia* (2021).

[44] H. Su et al. 2019. Perceptual Quality Assessment of 3d Point Clouds. In *2019 IEEE International Conference on Image Processing (ICIP)*.

[45] Shishir Subramanyam, Jie Li, Irene Viola, and Pablo Cesar. 2020. Comparing the Quality of Highly Realistic Digital Humans in 3DoF and 6DoF: A Volumetric Video Case Study. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*.

[46] Christian Szegedy, Sergey Ioffe, and Vincent Vanhoucke. 2016. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *CoRR* (2016).

[47] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro. 2017. Geometric distortion metrics for point cloud compression. In *2017 IEEE International Conference on Image Processing (ICIP)*.

[48] Dong Tian, Hideaki Ochimizu, Chen Feng, Robert Cohen, and Anthony Vetro. 2017. Geometric distortion metrics for point cloud compression. In *2017 IEEE International Conference on Image Processing (ICIP)*.

[49] Irene Viola and Pablo Cesar. 2020. A Reduced Reference Metric for Visual Quality Evaluation of Point Cloud Contents. *IEEE Signal Processing Letters* (2020).

[50] Irene Viola, Shishir Subramanyam, and Pablo Cesar. 2020. A Color-Based Objective Quality Metric for Point Cloud Contents. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*.

[51] Jianqiang Wang, Dandan Ding, Zhu Li, and Zhan Ma. 2020. Multiscale Point Cloud Geometry Compression.

[52] Jianqiang Wang, Hao Zhu, Zhan Ma, Tong Chen, Haojie Liu, and Qiu Shen. 2019. Learned Point Cloud Geometry Compression. *CoRR* (2019).

[53] V Zakharchenko. 2018. "Algorithm description of mpeg-pcc-tmc2". *ISO/IEC JTC1/SC29/WG11 MPEG2018/N17767* (Jul 2018).

[54] Juan Zhang et all. 2014. A subjective quality evaluation for 3D point cloud models. In *2014 International Conference on Audio, Language and Image Processing*.

5.7 Subjective Quality Evaluation of Point clouds With 3D Stereoscopic Visualization

Subjective Quality Evaluation of Point Clouds with 3D Stereoscopic Visualization

J. Prazeres, M. Pereira and A. M. G. Pinheiro 2022

IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 2022, pp. 2861-2865

DOI: 10.1109/ICIP46576.2022.9897937

SUBJECTIVE QUALITY EVALUATION OF POINT CLOUDS WITH 3D STEREOSCOPIC VISUALIZATION

João Prazeres, Manuela Pereira, António M. G. Pinheiro

Universidade da Beira Interior & Instituto de Telecomunicações, Covilha, Portugal

2022 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-9620-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIP46576.2022.9897937

ABSTRACT

In this paper, a subjective evaluation of static point clouds encoded with several codecs is described. Unlike other studies, a stereoscopic 3D display was used to visualize the 3D representation. A set of six point clouds were encoded using a set of state of the art point cloud coding solutions, notably the two MPEG codecs V-PCC and G-PCC, a deep learning solution RS-DLPCC that was the response to a call for evidence on point cloud coding of JPEG Pleno, and the popular DRACO codec. The results of this subjective quality evaluation using a 3D representation visualized in a stereoscopic display were compared with a previous subjective study that used the same content visualized in a 2D display. For that, the results of both tests were compared with the Pearson correlation, Spearman rank order correlation, the root mean square error and the outlier ratio. Moreover, the two subjective evaluation results were statistically analysed to seek for any statistical difference. The two subjective evaluations reveal a very high level of similarity.

***Index Terms*—** Point cloud quality, Subjective quality evaluation, Point cloud coding.

1. INTRODUCTION

The current technological needs is experiencing an increasing necessity for emergent 3D data formats. The representation of this data is of the utmost importance, since 3D content is represented for a huge amount of information. Most applications require efficient coding to provide the means of efficient processing, storage and transmission of 3D data. The point cloud technology is a 3D representation method, consisting on a set of Cartesian coordinates (x, y, z), with a list of attributes associated to each element, such as RGB component, reflective information, physical sensor information or normal vectors. Point clouds allow an extremely accurate representation of an object or scene, making them a very powerful representation model. They allow a visualization of the represented object or scene from any viewing position or distance, and a reliable 3D representation for relatively efficient computer based processing.

Several solutions for point cloud compression have been researched in the past, most notably the MPEG codecs V-PCC (Video Point Cloud Compression) and G-PCC (Geometry Point Cloud Compression) [1]. Recently, deep learning based solutions have been proposed for point cloud compression [2–5]. In response to a JPEG Call for evidence, a deep learning solution entitled Resolution Scalable Deep Learning Point Cloud Compression (RS-DLPCC) was also proposed [4]. Finally we also consider the Draco codec[1], developed by google as it has gained recently some popularity. This four coding solutions, VPCC intra, GPCC, RS-DLPCC and Draco are considered in this study.

Point cloud quality evaluation methodologies have been considered in multiple works recently. The most relevant aim point cloud visualization and are based on subjective evaluation. In [6] and [7], geometry only point clouds were considered and quality models were established. In [7] a 3D representation was considered, after a surface reconstruction algorithm was applied, and was concluded that using a 3D visualization and a 2D visualization did not change the evaluation.

Compression artifacts using prior encoding schemes are evaluated in [8–10]. Current efforts account for the range of high-performing codecs, such as the ones reported in [1, 11, 12]. In [13], a subjective evaluation using AR is proposed, whereas in [14], a VR environment is used to evaluate point clouds. In [15], a point cloud toolbox was created, in order to aid subjective testing in such environment. In [16], different resolutions and noise types were considered in a VR subjective evaluation. In [17] an evaluation on point cloud denoising algorithms is reported.

In this work a subjective evaluation using 3D representations of the point clouds visualized in a 3D stereoscopic display is reported. This is important because the 3D display provides a richer visualization of point clouds that might be likely to produce different subjective evaluation. Although the effects of the 3D visualization was analysed in [7], that work used geometry only point clouds. As texture information is likely to mask some geometric distortions [1], it is im-

Research funded by the Portuguese FCT-Fundação para a Ciência e Tecnologia under the project UIDB/EEA/50008/2020, PLive X-0017-LX-20, and by operation Centro-01-0145-FEDER-000019 - C4 - Centro de Competencias em Cloud Computing.

[1] https://github.com/google/draco

Table 1: Used parameters for G-PCC and V-PCC.

G-PCC						V-PCC					
Rate	R01	R02	R03	R04	R05	Rate	R01	R02	R03	R04	R05
QP	46	40	34	28	22	Geometry QP	36	32	28	20	16
pQS	0.25	0.5	0.75	0.857	0.9375	Texture QP	47	42	37	27	22
						Occupancy Map	4				2

Rate	R01	R03	R05
QP	7	9	10

Table 2: QP for draco codec.

portant to study the reliability of the typical evaluation protocol used in [1, 18]. Moreover, this paper does not use any surface reconstruction method, apart growing the dimension of the points to avoid transparency in the point clouds surface [1]. The results of the 3D point cloud subjective quality evaluation are then compared with the results of a 2D subjective evaluation [18], allowing to understand if there is a need of using 3D displays for future quality evaluations of point clouds.

2. DESCRIPTION OF THE USED CODECS

A short description of the used codecs can be found in [18]. A very short description is presented in the following together with the used codecs parameters, listed in tables 1 and 2.

V-PCC (Video Point cloud comparession) [19], projects the point cloud in a set of planes, encoding those projections in the 2D domain. Those projections contain texture, depth and an occupancy map. The projections are coded with the HEVC video codec.

G-PCC (Geometry Point Cloud Compression) [20] has two methods for point cloud compression. One is the octree based method, while the other is the trisoup, based on a surface reconstruction using triangular primitives, after deconstructing the model in an octree structure.

RS-DLPCC [2] uses deep learning to compress the point cloud geometry. A latent representation of a point cloud is computed by an autoencoder framework. The interlaced block creation makes the scalability feature possible. The point cloud is divided into superblocks, further divided by interlaced downsampling. This creates eight interlaced blocks for each defined superblock. The resulting blocks are then coded separately, enabling random access.

Draco is a codec developed by Google. The codec uses KD-Tree [21] in order to organize 3D data in an efficient way. The codec continuously split the point cloud from the center, modifying the axes on each direction.

3. EVALUATION METHODOLOGY

3.1. Point Cloud Data Selection

For this study, a set of point clouds was selected, all containing geometry and texture information. The set consisted of a frame selected from the soldier and longdress dynamic point clouds[2]. Frames 1300 and 0690 were selected from soldier and longdress sequences. The point clouds rhetorician abd guanyin were selected from the EPFL dataset. The romanoillamp and bumbameuboi point clouds were selected from the University of São Paulo Point Cloud Dataset[3]. These four point clouds represent cultural heritage. The selected point clouds are represented in figure 1. Considering previous studies [1], this database has a larger diversity, including not only human bodies, but also some objects with different geometric and textural characteristics. This testing database allows a better analysis of the codecs performance and improved generalization of the subjective evaluation protocol [18].

3.2. Data generation

A 360° rotation with steps of 1° was applied to each point cloud resulting in 12 second video sequences displayed at 30 FPS using PCL visualizer. They allow the point cloud visualization of every single angle.

To create the 3D depth perception, different views of the point cloud were created. Firstly, the point clouds were scaled in the X and Z axis, using PCL library. For the stereoscopic representation of the point clouds, the right and left representations of the point cloud are required, one for each eye. The left and the right representations are shifted 1.5° and -1.5° considering as reference the frontal view (as used in [18]).

Then, to ensure the point clouds were updated at the same time, they were aggregated. The original stimulus for the left view with the original stimulus for the right view, and the same for the coded stimulus. The final frames are composed of the left view distorted, left view reference and right view distorted, right view reference, for visualization of the 3D with the distorted point cloud on the left side. For the visualization with the reference on the left side, the reference and distorted were swapped. This is needed because two subjects doing tests in a row, will show the reference and the distorted swapped (reference appears once in the left, and once in the right side of the screen). To create the proper stereoscopic effect a translation also needs to be applied to each content, given by,

$$N_x = \frac{\frac{S_x}{2}}{\Delta}, \quad \text{with} \quad S_x = Z_{rec} \times \frac{IPD}{Z_v} \tag{1}$$

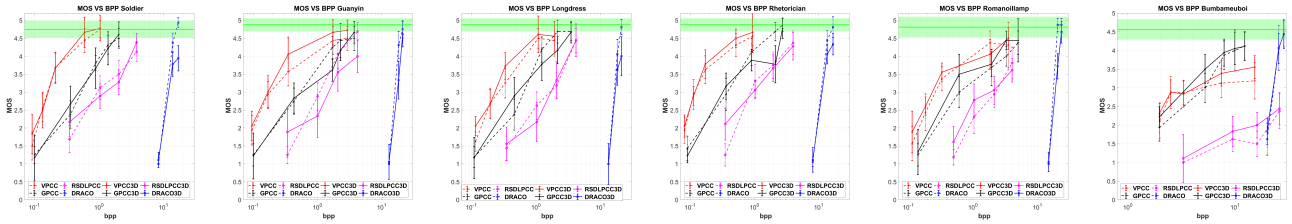
where Z_{rec} is the camera captured distance of the point cloud,Z_v is the visualization distance of the point cloud, IP

[2]https://jpeg.org/plenodb/

[3]http://uspaulopc.di.ubi.pt

(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Fig. 1: Point Cloud testing set.



(a) Soldier (b) Guanyin (c) Longdress (d) Rhetorician (e) Romanoillamp (f) Bumbameuboi

Fig. 2: MOS vs bpp with 95% confidence interval.

is the interpupillary distance, and Δ is the horizontal length corresponding to a sample in the camera (see table 3). The 3D volumes were located near the display plane to reduce the fatigue caused by the stereoscopic visualization.

The video sequences showing the rotation of the point clouds were created with FFMPEG, using the H.264 codec. It was ensured that no compression was applied by setting the CRF parameter to 0 and the q parameter to 0. To prevent issues with the RGB to YUV colorspace conversion the libx264rgb option was used for the video sequences creation.

The dimensions of the points for each point cloud used in the subjective evaluation is shown in 4. The point size is important to create continuous surfaces, avoiding perceptual effects caused by the transparency of the surfaces [6, 8].

A Double Stimulus Impairment Scale was used in this subjective evaluation. Hence, the subject visualizes both the reference and the coded point cloud side by side and was asked to evaluate the point cloud in a five-level rating scale (1 - very annoying, 2 - slightly annoying, 3 - annoying, 4 - perceptible, but not annoying, 5 - imperceptible).

Prior to the evaluation, the subjects were shown a sequence of four videos, with the redandblack point cloud (not included in the final test sequence) with four different levels of degradation, for familiarization with the artifacts created by the codecs. Additionally, hidden reference-reference pairs were included in the test sequence, resulting in a total of 108 pairs. The same content was never shown twice in a row. To avoid biases, half the subjects were shown videos with the reference on the right and distorted on the left, and vice-versa. All the tests were conducted in the subjective test laboratory of Image and Video Technology Group of Universidade da Beira Interior, using a 47 inch, FULL HD LG 47LA860V, with the test environment following the specifications in [22]. Both datasets are publicly available[4].

Table 3: Reconstruction Parameters.

Z_v	Z_{rec}	IDP	Δ
1.5 m	200	0.064	0.485

Table 4: Point size for each content.

	V-PCC					G-PCC				
Content	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05
Bumbameuboi	5	5	5	5	5	7	5	5	5	5
Guanyin	1	1	1	1	1	7	5	2	1	1
Longdress	1	1	1	1	1	7	5	2	1	1
Rhetorician	1	1	1	1	1	7	5	2	1	1
Romanoillamp	2	2	2	2	2	4	2	2	2	2
Soldier	1	1	1	1	1	7	5	2	1	1
	RS-DLPCC					Draco				
Bumbameuboi	-	23	16	10	10	10	-	6	-	5
Guanyin	-	7	5	3	1	8	-	3	-	2
Longdress	-	7	5	3	1	8	-	3	-	2
Rhetorician	-	7	5	3	1	8	-	3	-	2
Romanoillamp	-	9	4	2	2	9	-	5	-	2
Soldier	-	7	6	3	1	8	-	3	-	2

3.3. Subjective Evaluation

Table 6 shows the correlation values between the 3D test and the 2D test. Figure 3 shows the linear fitting values between

[4]http://webx.ubi.pt/∼pinheiro/icip2022pcdb.html

the two experiments. The high correlations between both tests suggest the subjects reacted similarly to both the 2D and 3D environments.

Table 5: Subject Information.

Males	Females	Overall	Age Span	Average age
15	3	18	22-47	28.35

Table 6: Statistical comparison between the 2D and 3D subjective tests.

No Fitting				Linear Fitting			
PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
0.958	0.937	0.094	0.529	0.958	0.937	0.085	0.529

The MOS results represented in figure 2 are very similar to those obtained in a 2D display reported in [18]. However, lower MOS values are obtained to the middle quality levels. After careful observation was concluded that the richer visual representation provided by the 3D display should be the cause, because it allows a better visualization of the coding artefacts.

A Kruskal-Wallis one way analysis followed by a multiple comparison [23] test was performed. The results for each content, each codec and for all are presented in table 7. It was concluded that there is no statistical significance between the subjective evaluation with the 3D and 2D visualizations (p value $>$ 0.05).

Table 7: Verification of statistical differences (Kruskal-Wallis).

Point Cloud	p values	Codec	p values
Bumbameuboi	0.809	VPCC	0.451
Guanyin	0.931	GPCC	0.594
Longdress	0.931	Draco	0.869
Rhetorician	0.890	RS-DLPCC	0.173
Romanoillamp	0.756	**All**	**0.7140**
Soldier	0.945		

3.4. Objective Evaluation

This quality evaluation study was also complemented with the analysis of the D1 and D2 [24], PSSIM [25] and PCQM [26] objective quality metrics. To evaluate the performance of the selected metrics the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Root-Mean Squared Error (RMSE) and the Outlier Ratio (OR) were computed as specified in [27]. Table 8 shows the results for both tests. The metric PCQM reveals the best results, confirming the good performance obtained for the 2D test. Fig. 4 shows the relation between MOS and the two best performing metrics.

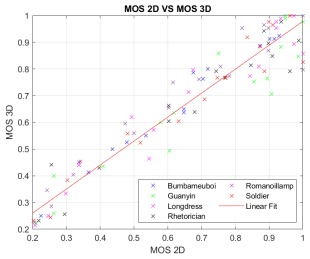
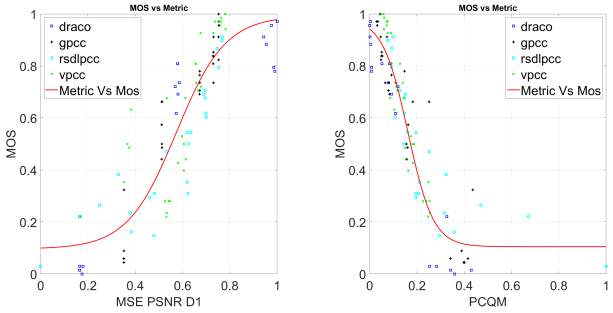


Fig. 3: Comparison between the MOS obtained using the 2D and the 3D representations, and respective linear fitting.



(a) MSE PSNR D1 (b) PCQM

Fig. 4: Relation between metrics and MOS, and Logistic fitting curve.

Table 8: Metrics performance using as ground truth the 2D and 3D subjective evaluations.

	3D Evaluation				2D Evaluation			
Metric	***PCC***	***SROCC***	***RMSE***	***OR***	***PCC***	***SROCC***	***RMSE***	***OR***
MSE PSNR D1	0.882	0.893	0.142	0.716	0.890	0.884	0.148	0.618
MSE PSNR D2 (Quadric 20)	0.855	0.851	0.157	0.657	0.851	0.847	0.169	0.608
MEAN PSSIM (Quadric 10)	0.871	0.866	0.148	0.686	0.866	0.863	0.162	0.627
PCQM	**0.934**	**0.924**	**0.108**	**0.637**	**0.944**	**0.928**	**0.106**	**0.480**

4. CONCLUSION

A quality subjective evaluation of point clouds using a 3D representation in a stereoscopic display is reported. Four codecs were used for the quality evaluation, V-PCC, G-PCC, RS-DLPCC and Draco codecs. Our study reveals that doing the subjective evaluation using a 3D visualization instead of a 2D visualization [1, 18] results in very similar results, that are highly correlated and do not reveal any statistical difference. However, subjects tend to give higher scores when visualizing the 3D stereoscopic representation instead of the 2D representation.

Finally the study was complemented with an analysis of objective metrics performance. The PCQM metric reveals a very good representation of the subjective results.

5. REFERENCES

[1] Stuart Perry et al., "Quality evaluation of static point clouds encoded using MPEG codecs," in *IEEE International Conference on Image Processing (ICIP)*, 2020.

[2] André F. R. Guarda et al, "Deep learning-based point cloud geometry coding with resolution scalability," in *IEEE 22nd Inter. Workshop MMSP*, 2020.

[3] André F. R. Guarda et al, "Adaptive deep learning-based point cloud geometry coding," *IEEE Journal of Selected Topics in Signal Processing*, 2021.

[4] André F. R. Guarda et al, "Point cloud coding: Adopting a deep learning-based approach," in *Picture Coding Symposium (PCS)*, 2019.

[5] Maurice Quach et al, "Learning convolutional transforms for lossy point cloud geometry compression," *IEEE International Conference on Image Processing (ICIP)*, 2019.

[6] E. Alexiou et al., "Point cloud subjective evaluation methodology based on 2D rendering," in *Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018.

[7] E. Alexiou et al., "Point cloud subjective evaluation methodology based on reconstructed surfaces," in *Applications of Digital Image Processing XLI*. International Society for Optics and Photonics, 2018, vol. 10752, SPIE.

[8] Luis A. da Silva Cruz et al, "Point cloud quality evaluation: Towards a definition for test conditions," in *Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019.

[9] Alireza Javaheri et al, "Subjective and objective quality evaluation of compressed point clouds," in *IEEE 19th Inter. Workshop MMSP*, 2017.

[10] S. Perry et al., "Study of subjective and objective quality evaluation of 3D point cloud data by the jpeg committee," *Electronic Imaging*, 2019.

[11] E. Alexiou et al., "A comprehensive study of the rate-distortion performance in MPEG point cloud compression," *APSIPA Trans. on Sign. and Infor. Proc.*, 2019.

[12] Su. Honglei et al, "Perceptual quality assessment of 3D point clouds," in *IEEE International Conference on Image Processing (ICIP)*, 2019.

[13] Evangelos Alexiou et al, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *IEEE 19th Inter. Workshop MMSP*, 2017.

[14] Shishir Subramanyam et al, "Comparing the quality of highly realistic digital humans in 3DoF and 6DoF: A volumetric video case study," in *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020.

[15] Evangelos Alexiou et al, "PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020.

[16] Juan Zhang et al, "A subjective quality evaluation for 3d point cloud models," in *International Conference on Audio, Language and Image Processing*, 2014.

[17] Alireza Javaheri et al, "Subjective and objective quality evaluation of 3D point cloud denoising algorithms," in *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2017.

[18] João Prazeres et al, "Quality analysis of point cloud coding solutions," in *Electronic Imaging Symposium*, 2022.

[19] MPEG 3DG, ," V-PCC Codec Test Model v8 ISO/IEC JTC1/SC29/WG11 W18884, Geneva, CH, October 2019.

[20] MPEG 3DG, ," G-PCC Codec Description v5 ISO/IEC JTC1/SC29/WG11 N18891, Geneva, CH, October 2019.

[21] O. Devillers and P.-M. Gandoin, "Geometric compression for interactive transmission," in *Proceedings Visualization 2000. VIS 2000 (Cat. No.00CH37145)*, 2000.

[22] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures,," Jan 2012.

[23] William H. Kruskal and W. Allen Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*.

[24] Dong. Tiang et al, "Geometric distortion metrics for point cloud compression," in *IEEE International Conference on Image Processing (ICIP)*, 2017.

[25] Evangelos Alexiou et al, "Towards a point cloud structural similarity metric," in *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2020.

[26] Gabriel Meynet et al, "PCQM: A full-reference quality metric for colored 3D point clouds," in *20th QoMEX*, 2020.

[27] ITU-T P.1401, "International telecommunication union,," in *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*, Jul 2012.

5.8 JPEG Pleno Call For Proposals Responses Quality Assessment

JPEG Pleno Call for Proposals Responses Quality Assessment
J. Prazeres, Z. Luo, A. M. G. Pinheiro, L. A. da Silva Cruz and S. Perry
ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5
DOI: 10.1109/ICASSP49357.2023.10094713

ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10094713

JPEG PLENO CALL FOR PROPOSALS RESPONSES QUALITY ASSESSMENT

Joao Prazeres[⋆] *Zhe Luo*[◇] *Antonio M. G. Pinheiro*[⋆] *Luis A. da Silva Cruz*[†] *Stuart Perry*[◇]

[⋆] Universidade da Beira Interior and Instituto de Telecomunicacoes, Portugal
[†]Universidade de Coimbra and Instituto de Telecomunicações, Portugal
[◇] University of Technology Sydney, Australia

ABSTRACT

In this paper, the quality evaluation of the responses to the Call for Proposals (CfP) of JPEG Pleno Point Cloud Coding is presented. Three responses to the CfP were evaluated together with the state of the art anchor codecs G-PCC and V-PCC from MPEG. The JPEG committee selected a set of eight point clouds that were encoded at different pre-established bitrates. For the subjective evaluation of the responses to the CfP, a set of video sequences were created where the reference and distorted decoded point clouds were rotated about their axes side by side. Furthermore, the objective quality metrics PCQM, PSNR D1, PSNR D2, PSNR Y and PSNR YUV were computed, and compared with the subjective evaluation results. This study revealed that the deep learning solutions outperformed G-PCC but were still below the performance of V-PCC regarding color representation. PCQM showed the best performance in predicting the compression quality.

Index Terms— Subjective evaluation, Objective Evaluation, Point Cloud, Deep-Learning

1. INTRODUCTION

Recently point clouds have become a hot topic in the research community, due to their wide range of applications. The point cloud data representation maps surfaces on a Cartesian coordinate system (x, y, z). Each mapped point might have a list of associated attributes, including RGB components, reflectance, physical sensor information, or normal vectors. This type of model can provide an accurate and detailed 3D representation of different objects or scenes. Point clouds can easily contain several million points, requiring efficient coding solutions for their storage and transmission.

The JPEG Committee has been working on coding standards for plenoptic data as part of its JPEG Pleno [1, 2] activity for a number of years. Plenoptic data in this context includes holography, light fields and point clouds, all of which are different representations of the plenoptic capture function. The scope of the JPEG Pleno Point Cloud activity is the development of standards for point cloud representation that not only involve efficient coding, but also support machine vision applications. During the 94th JPEG meeting, the JPEG Committee released a Final Call for Proposals (CfP) on JPEG Pleno Point Cloud Coding. This CfP addressed the first stage of the activity [3], which targeted a learning-based coding standard addressing human visualization and decompressed/reconstructed domain 3D processing and computer vision tasks.

Point cloud quality evaluation methodologies have been studied on numerous occasions. An evaluation of the MPEG standards for point clouds was previously reported [4]. Crowd sourcing was also employed in subjective evaluation in a recent work [5]. The two types of subjective evaluation reveal a very high level of statistical similarity. Recently, a subjective quality evaluation using 3D stereoscopic visualization was compared with 2D visualization revealing high correlations and no statistical difference [6]. Moreover, a subjective quality evaluation targeting machine-learning based coding solutions was reported [7]. Early in 2022 a quality assessment study was performed to support the JPEG Pleno point cloud coding CfP [8]. This study aimed to evaluate the current state of the art point cloud solutions, analyse the stability of the subjective quality assessment models and objective metrics performance.

To complement the subjective evaluation, several objective quality metrics have been proposed. The most common approaches mainly assess geometry and rely on Euclidean distances or projected errors along normal vectors [9]. In [10], color errors based on MSE and PSNR are applied in either the RGB or YCbCr color space, providing a relatively reliable metric using both geometry and color information. Other relevant works using both geometry and color information are presented in [11, 12]. A broad study of objective quality metrics was conducted in [13].

Three different codecs were submitted as responses to the CfP. These CfP responses are represented in this paper as T1 [14], T2 [15] and T3 [16]. These codecs are all based on deep learning, and are able to encode both geometry and color as requested by the CfP. In this article these codecs are compared against MPEG's G-PCC [17] and V-PCC [18]. Sections 2 and 3 describe the subjective and objective quality evaluations,

This work was funded by FCT/MCTES under the project UIDB/50008/2020 and project PLive X-0017-LX-20.

(a) *RWT130* (b) *RWT136* (c) *RWT246* (d) *RWT305*

(e) *RWT374* (f) *RWT395* (g) *RWT503b* (h) *RWT529*

Fig. 1: Point cloud test set.

respectively. Section 4 provides final conclusions.

2. SUBJECTIVE EVALUATION

2.1. Selected dataset

This experiment used the eight different point clouds shown in figure 1. This dataset contains meshes that were converted to point clouds following the methodology described in [19]. Table 1 shows the point cloud sparsity, colour gamut volume and mean curvature. The sparsity is defined as the average distance between each point and its 20 nearest neighbours, averaged over the entire point cloud. The point cloud generation methodology created a similarity of sparsity within point clouds in the test set which is a limitation of this test set. The colour gamut volume is defined as the volume of the Convex Hull of the distribution of colour points in the CIE 1976 LAB space divided by the volume of the CIE 1976 LAB colour space. The curvature values are computed using the method described in [20].

Table 1: Point cloud characteristics.

Point Cloud	Sparcity (K=20)	Colour Gamut Volume	Mean Curvature
RWT130	1.719	21%	0.137
RWT136	1.684	4%	0.145
RWT246	1.785	7%	0.104
RWT305	1.736	11%	0.102
RWT374	1.718	1%	0.121
RWT395	1.658	3%	0.163
RWT503b	1.662	5%	0.149
RWT529	1.713	2%	0.106

2.2. Subjective evaluation procedure

To prepare the data for the subjective evaluation, the dataset was encoded with the codecs submitted by the proponents and the anchors. The proponents had to target four different

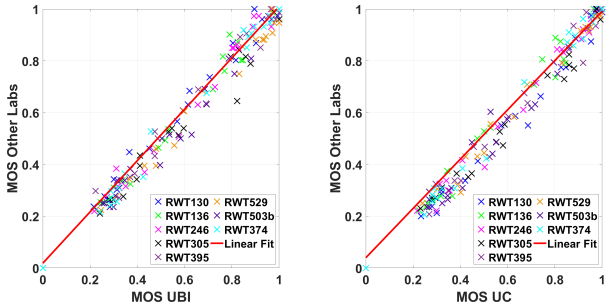


Fig. 2: Linear fitting between MOS obtained in different test laboratories.

Table 2: Testing laboratories information.

Display Information			
	UBI	UC	UTS
Monitor type	Eizo Color Edge CG318-4K	Eizo CG319X 4K HDR	Eizo CG318-4K
Resolution	4096×2160		
Distance	1.2 m	90 cm	1 m
Subject Information			
Males	12	11	6
Females	4	5	4
Average	27.35	32.9	33.9

Table 3: MOS statistical similarity between different laboratories.

	PCC	SROCC	RMSE	OR
UBI vs UC	0.987	0.977	0.055	0.357
UBI vs Other Labs	0.989	0.978	0.052	0.339
UC Vs Other Labs	0.989	0.982	0.052	0.333

bitrates selected by the JPEG Pleno Point Cloud committee, representing four different quality levels, ranging from low perceptual quality to high perceptual quality. After prior studies by the committee the bitrates $R1 = 0.1$ $R2 = 0.3$ $R3 = 1$ $R4 = 3$ bits per point (bpp) were selected, with a tolerance margin of +-10%.

Regarding video preparation, several point cloud views were captured using CloudCompare [1], each representing a 1° rotation. A complete rotation about the vertical axis was depicted, and the full sequence was rendered at 30 fps, thus resulting in 12 second videos. These video sequences were created with FFMPEG [2], using the H.264 codec [21].

The Constant Rate Factor (CRF) and q parameters were set to 0, so that no compression was applied. Furthermore, to prevent any RGB to YUV colorspace conversion the `libx264rgb` option was also used.

In some cases, the point size was changed to provide an improved visual representation. If holes appear in the point

[1] https://www.danielgm.net/cc/
[2] https://ffmpeg.org/

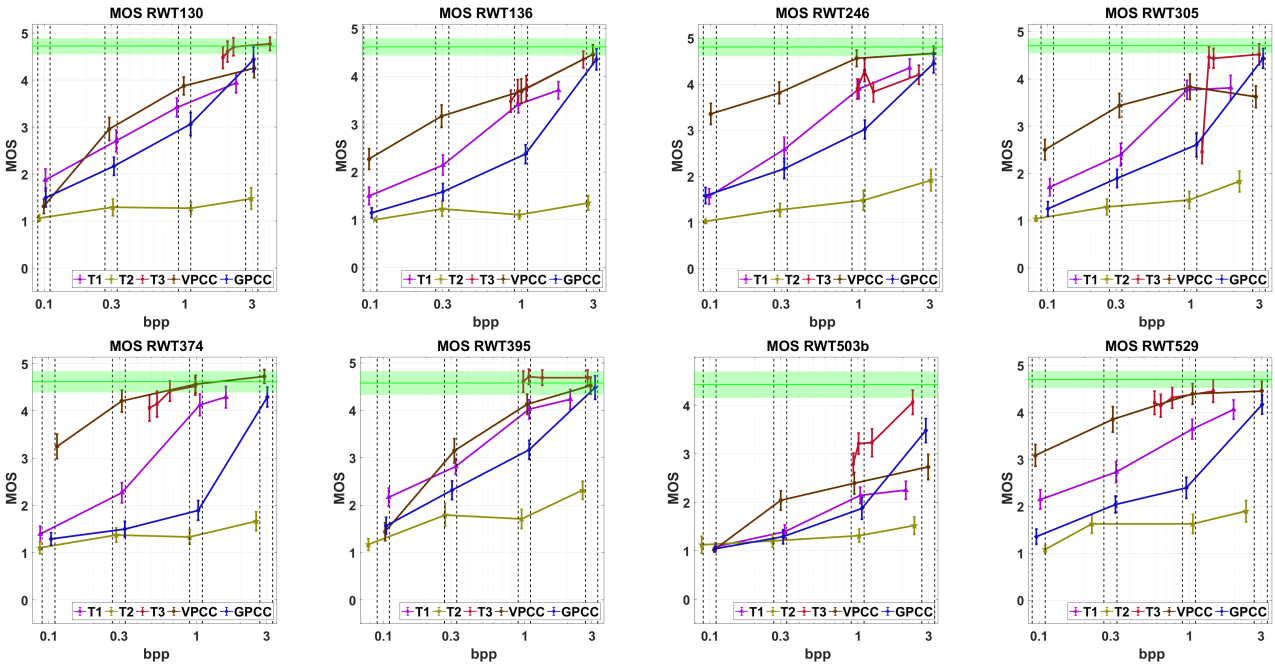


Fig. 3: MOS results for the tested point clouds. The error bars are 95% confidence intervals, while the green bars represent the 95% confidence intervals for the reference-reference stimuli. The vertical bars represent the tolerance margin for the bitrate (+-10%).

cloud the viewers will see the opposite part of the point cloud and that creates a negative quality perception [22, 23]. The change of point size is important to avoid that perceptual effect and create continuous surfaces for the point cloud under observation. This was achieved using CloudCompare, before the frame extraction process. For the test, the Double Stimulus Impairment Scale was used. In this method, the subject is shown both the reference and the coded point cloud, and was asked to evaluate the point cloud with a five-level rating scale (1 - very annoying, 2 - slightly annoying, 3 - annoying, 4 - perceptible, but not annoying, 5 - imperceptible). Prior to the evaluation, the subjects were shown a sequence of videos of the RWT53 point cloud, which was not included in the formal subjective test. The point cloud was encoded with the codecs under evaluation, for each chosen bitrate. This was done so that the subjects could familiarize themselves with the artefacts created by the codecs. Additionally, hidden reference-reference pairs were included in the test sequence, to help identify unusual behaviour in the evaluation. The same content was never shown twice in a row. To avoid biases, half the subjects were shown videos with the reference on the right and the coded content on the left, and the other half of the subjects were shown the vice-versa configuration. All the tests were conducted at Universidade da Beira Interior (UBI), Universidade de Coimbra (UC), both in Portugal, and University of Technology Sydney (UTS) in Australia, with the test environment following the specifications in [24]. Table 2 shows information about the equipment used in the subjective evaluation and information about the participants. Table 3 shows the correlation between sets of laboratories and Figure 2 shows the linear fitting between the results obtained. Additionally, a Kruskall-Wallis test was conducted, to check for statistical differences between the reported results. The p values obtained were above 0.05 ($p = 0.903$, for UBI vs UC, $p = 0.861$ for UBI vs Other Labs and $p = 0.718$ for UC vs Other Labs), revealing no statistical differences between the evaluations.

Figure 3 shows the subjective evaluation results. The bitrate for each tested content is computed as the ratio of the total number of bits of the encoded content divided by the number of input points in the encoded point cloud. It can be observed that the deep learning solutions T1 and T3 outperforms G-PCC. T1 is not able to reach the performance of V-PCC. T2 achieves comparable results in the lower bitrates, but cannot reach the performance of the other codecs as the rate increases. T3 could not attain the lower bitrates specified by the CfP.

3. OBJECTIVE EVALUATION

Subjective quality assessment provides a ground truth for the validation of objective quality metrics in the presence of the distortions produced by these codecs. In this paper, the performances of a selected set of objective quality metrics are

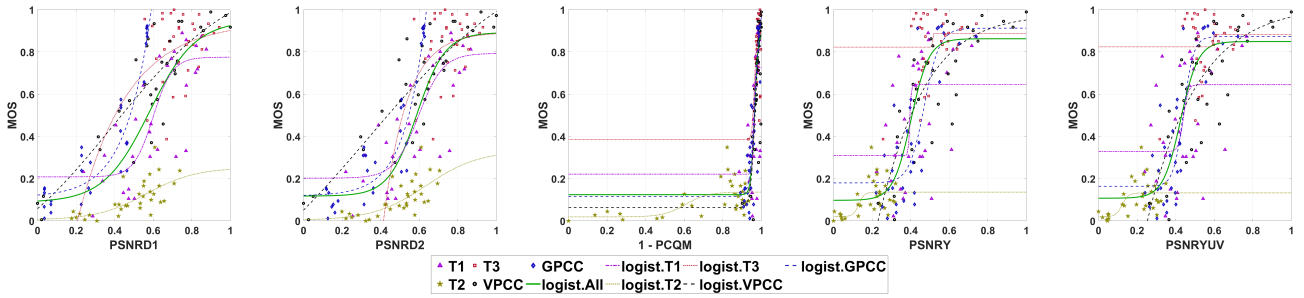


Fig. 4: MOS results plotted against objective metrics and the regression curve between each metric and the MOS values.

reported, notably for the PSNR D1 and PSNR D2 metrics [9], PSNR Y and PSNR YUV [25] and Point Cloud Quality Metric (PCQM) [10], by correlating their predicted MOS with the subjective MOS. These metrics were selected based on the JPEG Pleno Point Cloud Coding Common Training and Test Conditions document. The statistical measures proposed in [26] were computed for each of the above metrics, specifically the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Root-Mean Squared Error (RMSE) and the Outlier Ratio (OR). The predicted MOS for each of the objective metrics was computed applying a logistic fitting function on the objective scores. Table 4 reports the performance of the different metrics. PCQM was revealed to be the best performing metric, with $PCC = 0.873$ and $SROCC = 0.826$. It also reports the performance for each individual codec. For V-PCC, PSNR D1 and PSNR D2 show the best performance. Moreover, the PSNR D2 metric shows the best performance in evaluating G-PCC, T1 and T2, while PCQM achieves the best results when evaluating T3. Figure 4 shows the logistic fitting results.

Table 4: Metrics performance.

Global					
Metric	***Type***	***PCC***	***SROCC***	***RMSE***	***OR***
PSNR D1	FR_{GEO}	0.741	0.725	0.226	0.781
PSNR D2 (Quadric R =20)	FR_{GEO}	0.782	0.773	0.210	0.769
PCQM	$FR_{GEO+COL}$	**0.873**	**0.826**	**0.167**	**0.694**
PSNR Y	FR_{COL}	0.828	0.808	0.190	0.719
PSNR YUV	FR_{COL}	0.830	0.806	0.188	0.744
V-PCC					
PSNR D1	FR_{GEO}	**0.944**	**0.897**	**0.103**	**0.313**
PSNR D2 (Quadric R =20)	FR_{GEO}	**0.947**	**0.872**	**0.100**	**0.281**
PCQM	$FR_{GEO+COL}$	0.860	0.808	0.162	0.375
PSNR Y	FR_{COL}	0.822	0.687	0.178	0.531
PSNR YUV	FR_{COL}	0.865	0.8735	0.157	0.406
G-PCC					
PSNR D1	FR_{GEO}	0.876	0.796	0.132	0.344
PSNR D2 (Quadric R =20)	FR_{GEO}	**0.894**	**0.840**	**0.123**	**0.313**
PCQM	$FR_{GEO+COL}$	0.766	0.676	0.181	0.313
PSNR Y	FR_{COL}	0.824	0.790	0.156	0.500
PSNR YUV	FR_{COL}	0.781	0.741	0.171	0.563
T1					
PSNR D1	FR_{GEO}	0.797	0.768	0.166	0.594
PSNR D2 (Quadric R =20)	FR_{GEO}	**0.813**	**0.778**	**0.160**	**0.563**
PCQM	$FR_{GEO+COL}$	0.706	0.629	0.198	0.438
PSNR Y	FR_{COL}	0.546	0.480	0.229	0.719
PSNR YUV	FR_{COL}	0.484	0.452	0.239	0.813
T2					
PSNR D1	FR_{GEO}	0.688	0.751	0.061	0.469
PSNR D2 (Quadric R =20)	FR_{GEO}	**0.695**	**0.763**	**0.060**	**0.469**
PCQM	$FR_{GEO+COL}$	0.590	0.487	0.067	0.625
PSNR Y	FR_{COL}	0.614	0.500	0.066	0.625
PSNR YUV	FR_{COL}	0.595	0.522	0.067	0.625
T3					
PSNR D1	FR_{GEO}	0.209	0.068	0.151	0.719
PSNR D2 (Quadric R =20)	FR_{GEO}	0.254	0.117	0.149	0.688
PCQM	$FR_{GEO+COL}$	**0.540**	**0.277**	**0.130**	**0.531**
PSNR Y	FR_{COL}	0.339	0.085	0.146	0.656
PSNR YUV	FR_{COL}	0.272	0.082	0.148	0.656

4. DISCUSSION OF THE RESULTS

Analyzing the results from both Section 2 and 3, several conclusions can be drawn. The results from all laboratories were highly correlated, showing the robustness of the quality model, as observed in different studies [4, 8].

Analysing Table 4, it can be observed that PCQM is the best performing metric. Both full reference color metrics show good correlations between the metrics and MOS, higher than both full reference geometry only metrics. Table 4 also shows that the most common metrics fail to provide a reliable evaluation of the deep learning based codecs. For both G-PCC and V-PCC, the correlation values are high. When analysing the results for the deep learning based solutions, a significant drop in the metrics performance can be observed, mainly for the T3 solution. In this extreme case, the best correlation is a PCC of 0.540 and SROCC of 0.277 for the PCQM metric. Similar results have been reported in other studies [7, 27]. This reveals the need for research into new point cloud objective metrics that can accurately predict the quality of deep learning based solutions.

The deep learning based solutions are already revealing very close performance to V-PCC, although they are a novel approach for point cloud coding, A careful observation reveals that they already perform very well compressing geometry but they still need to be improved on the color domain.

5. REFERENCES

[1] Peter Schelkens et al, "JPEG Pleno: a standard framework for representing and signalling plenoptic modalities," Bellingham, 2018, vol. 10752 of *Proceedings of SPIE*, p. 107521P, SPIE-INT SOC OPTICAL ENGINEERING.

[2] Pekka Astola et al, "JPEG Pleno: Standardizing a coding framework and tools for plenoptic imaging modalities," *ITU Journal: ICT Discoveries*, vol. 3, Jun 2020.

[3] ISO/IEC JTC1/SC29/WG1, "Final call for proposals on JPEG Pleno Point Cloud Coding, Doc. WG1N100097," Jan 2022.

[4] Stuart Perry et al, "Quality evaluation of static point clouds encoded using MPEG Codecs," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020.

[5] Stuart Perry et al, "Comparison of remote subjective assessment strategies in the context of the JPEG Pleno point cloud activity," in *IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*, 2021.

[6] João Prazeres et al, "Subjective quality evaluation of point clouds with 3D stereoscopic visualization," in *IEEE International Conference on Image Processing (ICIP)*, 2022.

[7] Joao Prazeres et al, "Quality evaluation of machine learning-based point cloud coding solutions," in *Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis*, 2022, APCCPA '22.

[8] Stuart Perry et al, "Subjective and objective testing in support of the JPEG Pleno point cloud compression activity," *European Workshop on Visual Information Processing (EUVIP)*, September 2022.

[9] Dong Tian et al, "Geometric distortion metrics for point cloud compression," in *IEEE International Conference on Image Processing (ICIP)*, 2017.

[10] Gabriel Meynet et al, "PCQM: A full-reference quality metric for colored 3D point clouds," in *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020.

[11] Evangelos Alexiou and Touradj Ebrahimi, "Towards a point cloud structural similarity metric," in *IEEE ICME Workshops*, 2020.

[12] Irene Viola et al, "A color-based objective quality metric for point cloud contents," in *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020.

[13] Davi Lazzarotto, Evangelos Alexiou, and Touradj Ebrahimi, "Benchmarking of objective quality metrics for point cloud compression," in *IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*, 2021, pp. 1–6.

[14] ISO/IEC JTC1/SC29/WG1, "Doc. WG1M96010," Jan 2022.

[15] ISO/IEC JTC1/SC29/WG1, "Doc. WG1M96012," Jan 2022.

[16] ISO/IEC JTC1/SC29/WG1, "Doc. WG1M96095," Jan 2022.

[17] K. Mammou et al, "G-PCC codec description v2," *ISO/IEC JTC1/SC29/WG11 N18189*, Jan 2019.

[18] V Zakharchenko, ""algorithm description of mpeg-pcc-tmc2"," *ISO/IEC JTC1/SC29/WG11 MPEG2018/N17767*, Jul 2018.

[19] Davi Lazzarotto and Touradj Ebrahimi, "Sampling color and geometry point clouds from shapenet dataset," 2022.

[20] Zachary Taylor et al, "Multi-modal sensor calibration using a gradient orientation measure," *J. Field Robot.*, vol. 32, no. 5, pp. 675–695, aug 2015.

[21] ITU H.264, "Recommendation h.264," Aug 2021.

[22] Evangelos Alexiou et al, "Point cloud subjective evaluation methodology based on 2D rendering," in *Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018.

[23] L. A. da Silva Cruz et al., "Point cloud quality evaluation: Towards a definition for test conditions," in *Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019.

[24] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures,," Jan 2012.

[25] BT709 ITU-R BT.70, "Parameter values for the hdtv standards for production and international programme exchange,," Jun 2015.

[26] ITU-T P.1401, "International telecommunication union,," in *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*, Jul 2012.

[27] João Prazeres et al, "Quality analysis of point cloud coding solutions," in *Electronic Imaging Symposium*, 2022.

5.9 JPEG Pleno Learning-Based Point Cloud Coding: A Performance Analysis

JPEG Pleno Learning-Based Point Cloud Coding: A Performance Analysis
J. Prazeres, R. Rodrigues, M. Pereira and A. M. G. Pinheiro,
2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 1890-1894
DOI: 10.1109/ICIP49359.2023.10222278

JPEG PLENO LEARNING-BASED POINT CLOUD CODING: A PERFORMANCE ANALYSIS

Joao Prazeres, Rafael Rodrigues, Manuela Pereira, Antonio M. G. Pinheiro

Universidade da Beira Interior & Instituto de Telecomunicacoes, Covilha, Portugal

2023 IEEE International Conference on Image Processing (ICIP) | 978-1-7281-9835-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIP49359.2023.10222278

ABSTRACT

In this paper, a stability analysis of the JPEG Pleno Learning-based Point Cloud Coding Verification Model (VmUC) is performed. The codec is a deep learning-based solution that is able to compress both color and geometry. Three different training sessions were conducted using the default training set and cost function, and six point clouds were encoded/decoded with the resulting operating points for six target distortion/bitrate ratios. The VmUC performance was compared with the MPEG codecs V-PCC and G-PCC, considering three objective metrics, notably PSNR MSE D1, PSNR MSE D2, and PCQM. PSNR MSE D1 was also computed at each training epoch for the six decoded point clouds. It is concluded that the VmUC is able to outperform G-PCC and V-PCC in geometry encoding. However, it is outperformed by V-PCC in terms of color encoding, namely across all three training sessions. Furthermore, it is also shown that the codec does not present a high level of stability, changing its performance considerably with different training sessions.

Index Terms— Point cloud coding, deep learning-based codecs, objective evaluation

1. INTRODUCTION

Point clouds emerged as a very popular method for 3D data representation. This technology represents 3D data with a set of Cartesian coordinates (x, y, z), and each coordinate may have a list of attributes associated with it, such as RGB components, reflective information, physical sensor information, or normal vectors. Point clouds are able to provide extremely accurate representations of a given artifact, object, landscape, or building, resulting in a very powerful 3D representation model. However, depending on the level of detail, point clouds might contain an extremely high amount of data, making their transmission, storage, and processing difficult. Hence, methods for efficiently coding point cloud content are essential for the success of this format.

Two of the most well known coding solutions were proposed by MPEG, namely Video Point Cloud Compression (V-PCC) [1] and Geometry Point Cloud Compression (G-PCC) [2], for dynamic and static point clouds, respectively. Recently, a new class of deep learning (DL)-based coding of point clouds has gained some popularity [3–5] leading JPEG to launch a call for proposals on JPEG Pleno Learning-based Point cloud Coding [6]. The responses to this call lead to the development of the JPEG Pleno Point Cloud Coding Verification Model, which is the aim of the current study. The evaluation of compression solutions requires reliable subjective and objective quality models. As subjective evaluation is time-consuming and typically requires careful planning and design, quality evaluation metrics are quite important for coding method developers.

Several studies on point cloud coding quality evaluation allowed for the analysis of a set of point cloud metrics [7–11] and to study its reliability. Based on these studies and also on the popularity of the metrics, the following ones were selected for this study: The full reference geometry only metrics point-to-point (PSNR MSE D1) and point-to-plane (PSNR MSE D2) [12], and the full reference joint metric PCQM [13]. Other relevant metrics, not considered in this study, have been proposed, like Point to Distribution [14], PointSSIM [15], or GraphSIM [16]. When learning-based methods are considered, the performance of the final learned model may vary because of the stochastic nature of the learning process, even when training conditions remain unchanged. An analysis was conducted earlier [17], revealing that some solutions may result in different quality levels of performance for different training sessions with the same conditions.

This paper aims to evaluate the stability of the Learning-based Point Cloud Coding Verification Model (VmUC) [18]. Three training sessions were conducted with the exact same conditions as in previous studies [17]. At each learning epoch, the PSNR MSE D1 metrics was computed in order to evaluate the metric variation across training sessions. These metrics were selected as they have shown good correlation with subjective evaluation conducted with DL-based solutions [10, 11, 19]. Furthermore, the final epochs of each method are compared with the MPEG codecs. This study allows us to understand how quality evolves during the training process. Furthermore, it shed insight on how final performance of the training sessions, showing how sensitive the codec is to different training.

In the following, a short codec description of the JPEG Pleno Learning-based point cloud coding verification model

Research funded by the Portuguese FCT-Fundação para a Ciência e Tecnologia under the project UIDB/50008/2020, PLive X-0017-LX-20, and by operation Centro-01-0145-FEDER-000019 - C4 - Centro de Competencias em Cloud Computing.

is given, followed by the experimental description and conclusions.

2. CODEC DESCRIPTION

2.1. JPEG Pleno Learning-based Point Cloud Coding Verification Model

The developed VmUC [18] is currently a DL-based solution that jointly encodes geometry and color. At the encoder side, the point cloud is partitioned into 3D blocks of fixed size, which are encoded separately. Each voxel within these blocks has four channels, carrying a binary occupancy value and normalized RGB values.

Before encoding, blocks are down-sampled, depending on the characteristics of the point cloud, such as their sparsity. Block encoding is done with an end-to-end DL coding model. At the first stage, an autoencoder (AE) combines 3D convolutional layers and Inception-Residual Blocks (IRB) [20]. The resulting latent representation is then quantized, with a given QS (quantization step), and entropy coded. Also, the AE latents are fed to a variational autoencoder (VAE), which allows to capture structural information still in the latent representation. The resulting VAE latents are used as hyperpriors for the AE entropy bottleneck, and also entropy coded.

The decoder architecture mirrors the encoder, using 3D convolutional layers and the IRB, and outputs blocks with values ranging from 0 to 1, representing the probability of voxel occupancy, as well as the RGB values. Probability values are transformed into binary values using an *optimized Top-k* binarization algorithm. Only the *k* voxels with the largest probabilities are selected as points, with *k* defined as where N_{input} represents the number of points on the reference block, and β is a factor defined via model optimization, by maximizing geometry and texture quality metrics, such as PSNR MSE D1 and PSNR MSE YUV.

2.2. JPEG Pleno Learning-based Point Cloud Coding Verification Model Training

The model is trained by minimizing a loss function which considers the distortion of each decoded block, comparing it to the input block, as well as the estimated coding rate. The loss function is given by $J = D + \lambda R$, involving the Lagrangian multiplier λ, which controls the rate distortion trade-off. R is the bitrate and D is the distortion, defined as: $D = (1-\omega) \times D_{Geo} + \omega \times D_{Col}$, where D_{Geo} is the geometry distortion and D_{Col} is the color distortion. The weight ω defines the balance between color and geometry, and was established as 0.5 [18]. The geometry distortion is defined as the average voxel level distortion. The color distortion is defined as a voxel-wise mean squared error between the RGB values of the occupied voxel in the input block and decoded block. The coding rate is estimated during training as the entropy of the AE and VAE latent representations.

The model is trained with a selection of the static PCs listed in the JPEG Pleno PCC Common Training and Test Conditions (CTTC) [21]. The point clouds were downsampled to a lower precision according to their sparsity and partitioned into blocks of $64 \times 64 \times 64$. The blocks with less than 500 filled voxels were removed. An early stopping of the training process was applied in order to prevent over-fitting. Training is stopped if the validation loss does not decrease after 5 epochs. Six working points were considered for this study, using λ = 0.000125, 0.00025, 0.0005,0.004, 0.002 and 0.001.

3. EXPERIMENT DESCRIPTION

3.1. Dataset Selection

The dataset was chosen in concordance with previous studies conducted on DL-based codecs [10, 11, 17]. The dataset contains six point clouds, depicting three objects: frame 1300 of the dynamic point cloud *Longdress*, available at the JPEG Pleno Database; the point cloud *Guanyin* from the EPFL Database; and *Romanoillamp* and three landscapes: the point clouds *Ramos*, *Citiusp* and *IpanemaCut*, from the São Paulo University database, depicted in Fig. 1.

3.2. Performance Stability

A study of the influence of the training model described in section 2.2 on the performance of the VmUC codec is established.

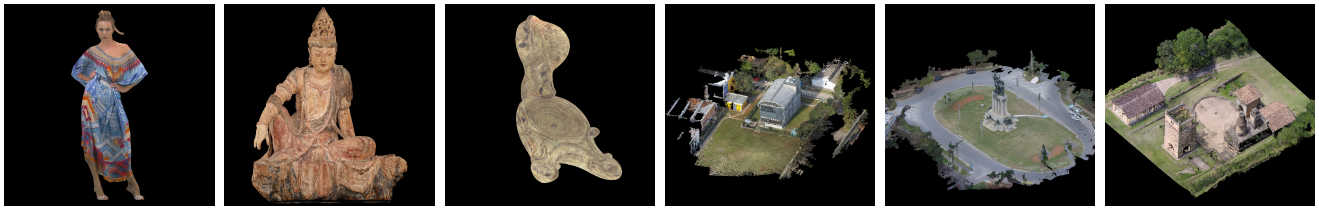
For this study, the codec was trained three times. The λ values described above were used, and the model was also trained using the early stopping mechanism. At each training

Table 1: Coding parameters for G-PCC and V-PCC.

	G-PCC						V-PCC					
Content	**Rate**	**R01**	**R02**	**R03**	**R04**	**R05**	**Rate**	**R01**	**R02**	**R03**	**R04**	**R05**
Longdress	QP	41	34	31	29	28	**Geometry QP**	32	28	24	22	18
	pQs	0.29	0.325	0.41	0.51	0.61	**Texture QP**	42	37	32	28	15
Guanyin	QP	41	34	32	30	28	**Geometry QP**	32	28	24	22	18
	pQs	0.29	0.355	0.51	0.6355	0.79	**Texture QP**	42	37	32	28	15
Romanoillamp	QP	41	38	35	33	30	**Geometry QP**	32	28	24	22	18
	pQs	0.24	0.33	0.47	0.61	0.7175	**Texture QP**	42	37	32	28	15
Citusp	QP	41	34	32	30	28	**Geometry QP**	32	28	24	22	18
	pQs	0.29	0.355	0.51	0.6355	0.79	**Texture QP**	42	37	32	28	15
Ramos	QP	41	34	32	28	26	**Geometry QP**	32	28	24	22	18
	pQs	0.31	0.355	0.48	0.5525	0.64	**Texture QP**	42	37	32	28	15
IpanemaCut	QP	39	36	34	30	28	**Geometry QP**	32	28	24	22	18
	pQs	0.315	0.45	0.62	0.70	0.73	**Texture QP**	42	37	32	28	15
							Occupancy Map	4				

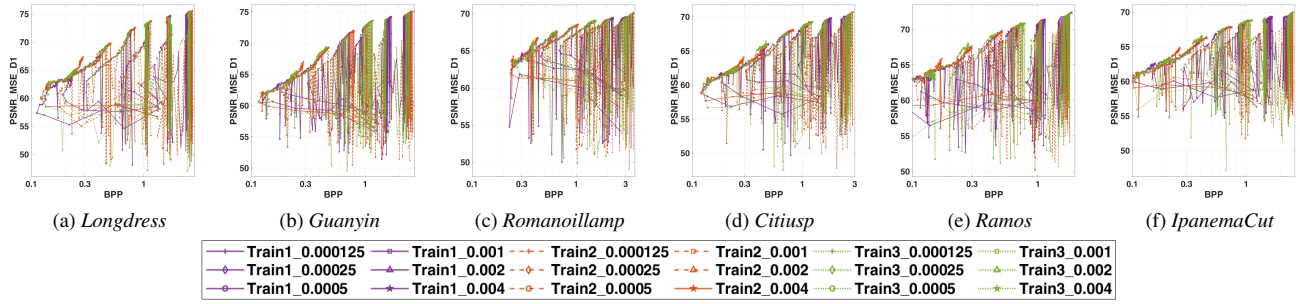
Table 2: BD-Metrics and BD-Rate using G-PCC as a reference.

	PSNR MSE D1		**PSNR MSE D2**		**PCQM**	
Codecs	**BD-PSNR**	**BD-Rate**	**BD-PSNR**	**BD-Rate**	**BD-Metric**	**BD-Rate**
VmUC Train 1	2.908	-26.302	2.204	-18.773	-1.746E-03	67.724
VmUC Train 2	3.369	-56.292	2.876	-51.759	-1.705E-03	48.335
VmUC Train 3	3.149	-56.964	2.660	-46.735	-2.240E-03	60.528
VmUC Default	4.359	-72.885	4.235	-66.397	-1.603E-03	48.193
V-PCC	3.128	-60.617	1.695	-26.307	1.406E-03	-27.885



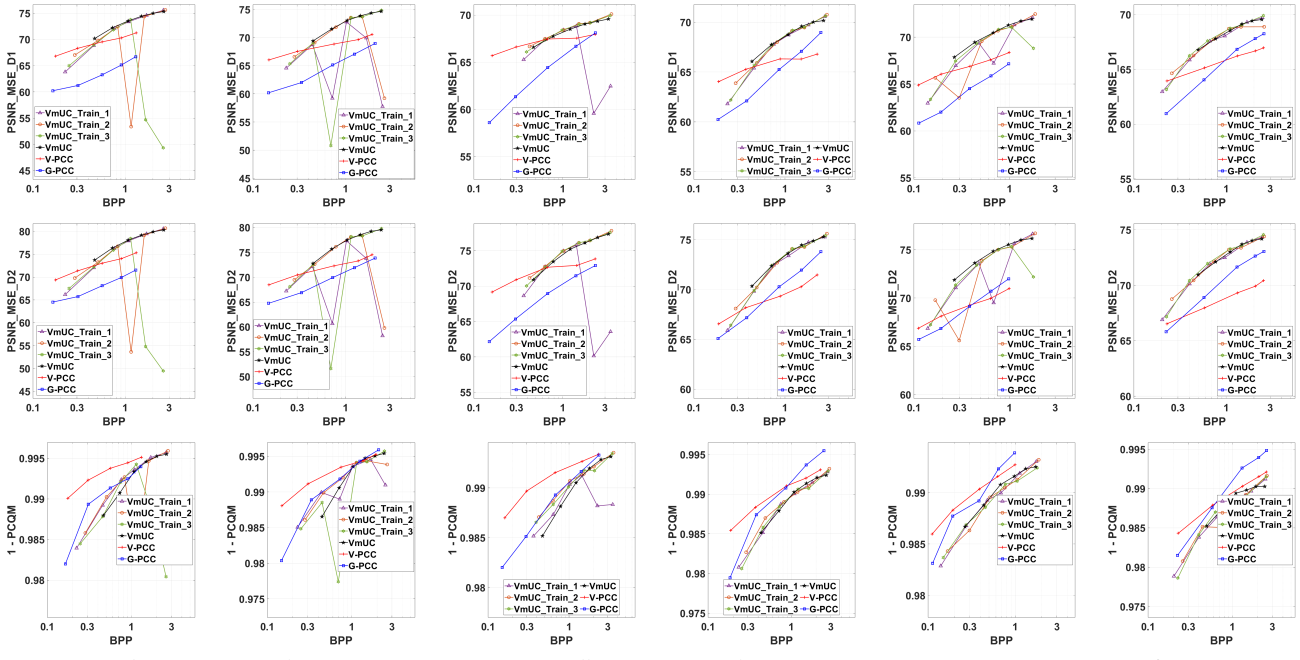
(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *Ramos* (f) *IpanemaCut*

Fig. 1: Point cloud dataset, in concordance with previous studies [17].



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *Ramos* (f) *IpanemaCut*

Fig. 2: PSNR MSE D1 vs. bpp plots for the VmUC at each training epoch.



(a) *Longdress* (b) *Guanyin* (c) *Romanoillamp* (d) *Citiusp* (e) *Ramos* (f) *IpanemaCut*

Fig. 3: Plots of the performance metrics for the default VmUC, the three trained models, and for the two MPEG anchors. The x axis represents the bitrate, and the y axis represents the metric.

epoch, a set of six point clouds were tested using the defined codec and the metrics PSNR MSE D1 and PCQM were com-

puted using the resulting decoded point cloud. Figure 2 shows the metric values for each epoch. For the lower bitrates, a continuous increase in quality is observed. However, for higher bitrates, a high level of instability can be observed across the training epochs, and all metrics show very erratic behavior, reaching very low and very high values between epochs. This might be caused by overfitting or a lack of suitable training data. Nevertheless, this is a very undesirable behavior that can cause bad coding performance at higher bitrates for some content.

3.3. Comparison with the MPEG anchors

Furthermore, an analysis comparing the codecs working points (resulting from each training procedure) with the MPEG anchors V-PCC and G-PCC was conducted. The test point clouds were encoded with bitrates as close as possible to the ones defined by VmUC, using the codec parameters defined in table 1.

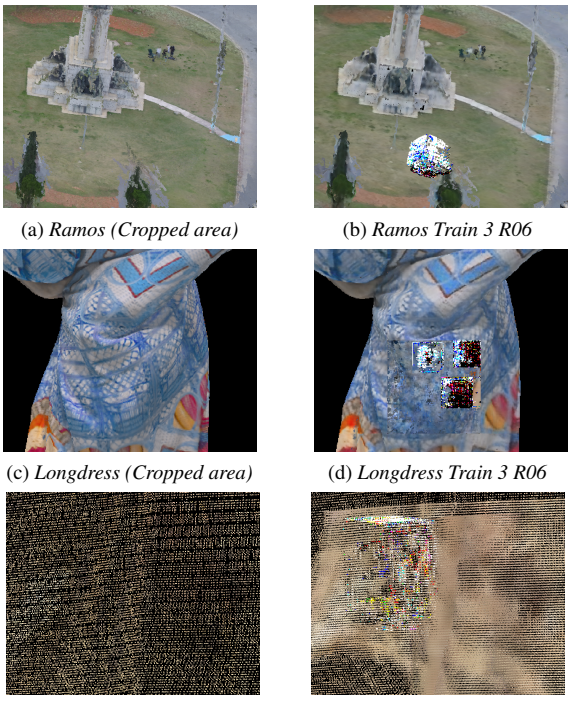
Figure 3 shows the plots of the metrics PSNR MSE D1, PSNR MSE D2 [12] and PCQM [13] computed for the decoded test point clouds. The VMuC results in three different plots, resulting from the three training sessions. The plots also show the results for the VmUC with the working points available in the default implementation [1]. For the *Longdress* point cloud, the second training always produces a drop on the middle bitrate, and the third training produces a really bad quality in the highest bitrate. The same is observed in the highest bitrate of the *Guanyin* and *Romanoillamp* contents for the second training as well. The *Ramos* point cloud produces some erratic behavior for the PSNR MSE D1 and PSNR MSE D2 metrics. Nevertheless, this unstable behavior is not observed in metrics like PCQM, which has been found to be more reliable [11]. In the future, subjective testing should be performed to evaluate these results and determine the best metric in this context. In the three point clouds representing landscapes, the performance of the codec was quite stable for PCQM.

Table 2 shows the BD rates and BD metrics [22] relative to the G-PCC. While the geometry metrics show a gain for the default and the three training cases, the PCQM reveals a slightly lower performance when compared with the G-PCC. Moreover, the geometry metrics also reveal the change in performance already observed.

3.4. Visual examples

Figure 4 shows three visual examples of high rates that created the unusual bad quality for high rates in some of the training sessions of the VmUC. It can be observed that some blocks were not reconstructed, in the higher bit rate of train 3 of *Ramos* and *Longdress* point clouds. The second highest rate of train 1 of *Romanoillamp* point cloud crop is also shown, where some visible distortions can be observed.

[1] https://gitlab.com/wg1/jpeg-pleno/jpeg-pleno-pc-vm



(a) *Ramos (Cropped area)* (b) *Ramos Train 3 R06*

(c) *Longdress (Cropped area)* (d) *Longdress Train 3 R06*

(e) *Romanoillamp (Cropped area)* (f) *Romanoillamp Train 1 R05*

Fig. 4: Cropped areas of selected decoded point clouds.

Moreover, the codec tends to create more points in the decoded point cloud, as the original point cloud is somehow sparse.

4. CONCLUSION

An analysis of the performance considering different training sessions of the JPEG Verification Model is presented. It was observed that, in general, the VMuC produces a better bitrate Distortion relation than the G-PCC but typically cannot reach the performance of V-PCC when comparing both color and geometry. The performance variations for each training session, where in some cases higher bitrates result in lower quality, reveal a current limitation that is typical of the DL-based codecs. This tends to happen at higher bitrates, where the codecs performance is stretched. Furthermore, it is important to emphasize that it is very likely that most applications will use the higher bitrates, as the lower ones produce high distortions. A possible solution is to have multiple profiles for each bitrate, and if the quality is not satisfactory, another profile can be tested. If better performance results, then the new profile needs to be used. Eventually, if the profile is not a default, the decoder settings also need to be included in the bitstream, typically in the metadata. At the time of the definition of this paper, JPEG is planning to rebuild its Verification Model architecture to provide an entirely independent compression of the point cloud geometry and color.

5. REFERENCES

[1] MPEG 3DG, ," V-PCC Codec Test Model v8 ISO/IEC JTC1/SC29/WG11 W18884, Geneva, CH, October 2019.

[2] MPEG 3DG, ," G-PCC Codec Description v5 ISO/IEC JTC1/SC29/WG11 N18891, Geneva, CH, October 2019.

[3] Jianqiang Wang et al, "Multiscale point cloud geometry compression," in *Data Compression Conference DCC*. IEEE, 2021, pp. 73–82.

[4] Maurice Quach et al, "Improved deep point cloud geometry compression," in *International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2020, pp. 1–6.

[5] André FR Guarda et al, "Deep learning-based point cloud geometry coding with resolution scalability," in *International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2020, pp. 1–6.

[6] ISO/IEC JTC1/SC29/WG1, "Final call for proposals on JPEG Pleno Point Cloud Coding, Doc. WG1N100097," Jan 2022.

[7] Stuart Perry et al., "Quality evaluation of static point clouds encoded using MPEG codecs," in *IEEE International Conference on Image Processing (ICIP)*, 2020.

[8] Stuart Perry et al, "Comparison of remote subjective assessment strategies in the context of the JPEG Pleno point cloud activity," in *International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2021, pp. 1–6.

[9] Davi Lazzarotto et al, "Benchmarking of objective quality metrics for point cloud compression," in *International Workshop on Multimedia Signal Processing (MMSP)*, 2021, pp. 1–6.

[10] João Prazeres et al, "Quality analysis of point cloud coding solutions," in *Electronic Imaging Symposium*, 2022.

[11] Joao Prazeres et al, "Quality evaluation of machine learning-based point cloud coding solutions," in *Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis*, 2022, APCCPA '22.

[12] Dong Tian et al, "Geometric distortion metrics for point cloud compression," in *IEEE International Conference on Image Processing (ICIP)*, 2017.

[13] Gabriel Meynet et al, "PCQM: A full-reference quality metric for colored 3d point clouds," in *International Conference on Quality of Multimedia Experience (QoMEX)*, 2020.

[14] Alireza Javaheri et al, "A point-to-distribution joint geometry and color metric for point cloud quality assessment," in *International Workshop on Multimedia Signal Processing (MMSP)*, 2021, pp. 1–6.

[15] Evangelos Alexiou et al, "Towards a point cloud structural similarity metric," in *International Conference on Multimedia Expo Workshops (ICMEW)*, 2020.

[16] Qi Yang et al, "Inferring point cloud quality via graph similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3015–3029, 2022.

[17] João Prazeres et al, "On the stability of point cloud machine learning based coding," in *European Workshop on Visual Information Processing (EUVIP)*, 2022.

[18] ISO/IEC JTC1/SC29/WG1, "IT/IST/IP Leiria response to the call for proposals on jpeg pleno point cloud coding v1.1 , Doc. WG1M96005," Jan 2022.

[19] João Prazeres et al, "Subjective quality evaluation of point clouds with 3D stereoscopic visualization," in *IEEE International Conference on Image Processing (ICIP)*, 2022.

[20] Christian Szegedy et al, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. 2017, AAAI'17, p. 4278–4284, AAAI Press.

[21] ISO/IEC JTC1/SC29/WG1, "JPEG Pleno Point Cloud Coding common training and test conditions v1.1 , Doc. WG1N100112," Jan 2022.

[22] Gisle Bjontegaard, "Calculation of average psnr differences between rd-curves," in *ITU - Telecommunications Standardization Sector: VCEG-M33*, March 2001.

5.10 Subjective Quality Evaluation Of Point Clouds Using a Head-Mounted Display

J. Prazeres, R. Rodrigues, M. Pereira and A. M. G. Pinheiro, ”Subjective Quality Evaluation of Point Clouds Using a Head-Mounted Display,”
ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5,
doi: 10.1109/ICASSP49660.2025.10889051.

ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10889051

Subjective Quality Evaluation of Point Clouds Using a Head-Mounted Display

Joao Prazeres*, Rafael Rodrigues*, Manuela Pereira†, Antonio M. G. Pinheiro*
*Universidade da Beira Interior & Instituto de Telecomunicacoes, Covilha, Portugal
†Universidade da Beira Interior & NOVA LINCS, Covilha, Portugal

***Abstract*—Preparing subjective evaluations for point cloud quality assessment using 2D or 3D planar displays typically requires rendering uncompressed videos, which demands large volumes of data storage and processing, particularly at high resolutions. This paper describes a subjective quality evaluation of point clouds using a head-mounted display (HMD), which allows the direct 3D content depiction and thus eliminates the need for video rendering. The test dataset included six static point clouds encoded with the MPEG codecs V-PCC and G-PCC, the learning-based codec RS-DLPCC, and the Draco codec. The obtained Mean Opinion Scores (MOS) were compared with two other studies using the same dataset but with different displays, notably a 2D display and a 3D stereoscopic display. The reported results indicate that the three studies are highly correlated, with Pearson and Spearman correlation coefficients above 0.9. Moreover, a multi-way ANOVA shows that changing the display does not have a significant effect on the obtained MOS.**

***Index Terms*—Point cloud coding, Learning-based codecs, Subjective evaluation**

I. Introduction

Point clouds are a popular 3D representation method used in various fields like virtual reality (VR), augmented reality (AR), and 3D printing. They consist of points in a Cartesian coordinate frame (x, y, z), usually with attributes (i.e. RGB values) associated with each coordinate. However, efficient compression techniques and quality models are needed, as point clouds typically contain large amounts of data.

This study presents a subjective quality assessment of colored point clouds under coding distortions conducted using a head-mounted display (HMD). The results of this subjective experiment are compared with two others performed employing a two-dimensional (2D) display [1] and three-dimensional (3D) stereoscopic visualization [2]. Up to our knowledge, this is the first study that makes such comparison, allowing to understand the reliability of the developed methodology.

Alexiou *et al.* introduced PointXR [3], a software toolkit designed for the visualization and subjective evaluation of point clouds in VR environments. Viola *et al.* [4] conducted a subjective quality evaluation that compared different degrees of freedom, when evaluating dynamic point clouds with a HMD. Zhou *et al.* [5] conducted a subjective quality evaluation using an HMD and an eye tracker for dynamic point clouds. Wu *et al.* [6] conducted an evaluation of V-PCC using a HMD. Both evaluations were conducted with 6 degrees of freed (DoF).

Multiple studies have developed quality models for encoding methods that focus exclusively on geometry [7]–[10]. Perry *et al.* presents an assessment of the perceived quality of MPEG Point Cloud codecs using a 2D display [11]. An initial investigation was conducted to assess the quality of point cloud coding using learning-based techniques [12]. To allow an effective comparison this study uses the same testing dataset of the prior studies [1], [2], [13], obtained with the MPEG codecs V-PCC and G-PCC [14], RS-DLPCC [15]), and Draco[1].

Using HMDs for subjective quality assessment has many potential benefits. When preparing subjective tests for point cloud quality evaluation using 2D displays, it is common to prepare lossless high-definition videos (preferably 4K) of the point clouds rotating around a given axis, which could lead to 400GB+ videos. The requirement for lossless is caused by the fact that video compression may introduce additional undesirable artifacts. HMDs directly render the point cloud, thus eliminating the need for video rendering.

This paper main contributions are: 1) A protocol to use HMDs for subjective evaluation is defined, that provides an immersive visualization. 2) A comparison between the results obtained with the current study and those obtained with typical methodologies for point cloud subjective quality assessment using 2D displays.

II. Evaluation Methodology

This section provides details on the used dataset, the considered point cloud codecs, and the quality assessment model.

A. Point Cloud Data Selection

For this study, a set of six point clouds was selected (fig.1), all containing geometry and texture information. The set consisted of frames 1300 from the *Longdress* and 690 from the *Soldier* dynamic point clouds[2], the static point clouds *Rhetorician* and *Guanyin*, from the EPFL dataset, and *Romanoillamp* and *Bumbameuboi*, from the University of Sao Paulo Point Cloud Dataset[3]. The first two represent a human figure, and the last four represent cultural heritage, providing diversity of geometrical, textural, and point density characteristics within

This work is funded by FCT/MECI through national funds and when applicable co-funded EU funds under UID/50008: Instituto de Telecomunicações.

[1]https://google.github.io/draco/
[2]https://jpeg.org/plenodb/
[3]http://uspaulopc.di.ubi.pt

(a) *Longdress* (b) *Soldier* (c) *Rhetorician* (d) *Guanyin* (e) *Romanoillamp* (f) *Bumbameuboi*

Fig. 1: Point Cloud testing set.

the dataset [2]. The coding distortions were created with 4 different codecs. V-PCC [14] (using HEVC as a video codec) and G-PCC (using Octree for geometry and Prediction-plus-Lifting for texture) [14] codecs were selected as they are widely used state-of-the-art solutions. RS-DLPCC [15] is a learning-based solution with a scalability property, and Draco was chosen because of its popularity. The authors would also like to acknowledge the authors of [15] for providing decoded data from RS-DLPCC for the reported experiment. In total, a collection of 108 point clouds were evaluated in the subjective evaluation (reference point clouds included).

A brief description of the used codecs may be found in the papers related to our previous studies [1], [2], as well as the coding parameters.

B. Data Generation and Experimental Setup

The quality evaluations using 2D [1] and 3D [2] displays followed the Double-Stimulus Impairment Scale (DSIS) method [16], while the evaluation described in this paper followed an approach where the distorted and reference content can be alternately visualized in the HMD to allow an easy comparison, similar to a flicker test.

To improve the visual representation, the size of the points was changed according to table I. This ensures that the subjects view the opposite or inner section of the point cloud, as the point cloud's surface might include some transparency or empty areas, resulting in a highly subpar perception [7], [10]. This manipulation is important to avoid this perceptual effect by creating continuous surfaces. However, it was ensured that this manipulation did not mask compression artifacts.

Prior to the evaluation, the participants were shown a stimulus that showed a point cloud that wasn't part of the testing set, representing various degrees of distortion generated by the assessed codecs. This was conducted to make subjects familiar with the distortion artifacts produced by the codecs. The participants then used a five-level rating system (1: very annoying, 2: slightly annoying, 3: annoying, 4: perceptible but not annoying, and 5: imperceptible) to evaluate the quality of the distorted point cloud.

The visualization of sequentially distorted representations of the same content was avoided. Videos were shown to half the individuals with the reference on the right and the distorted information on the left, and vice versa, to prevent biases. Hidden reference-reference pairings were incorporated into the test sequence as well. In total, 18 naive subjects participated in the study. All subjects were students of the university and have passed a Snellen visual acuity test and Ishihara color vision test prior to the experiment. All sessions were conducted in the Universidade da Beira Interior Multimedia Quality Laboratory, and the evaluation environment was adjusted in accordance with ITU-R BT500-15 [16]. The point clouds used in the HMD evaluations were displayed using an HTC Vive Pro Headset with a refresh rate of 90 Hz, a field of view of 110 degrees, and a resolution of 2880 × 1600 (1440 × 1600 per eye). Information about the resolutions of the previous experiments can be found in our previous studies [1], [2].

To conduct the experiment, the Unity[4] software was used with the Pcx point cloud importer library[5], which allows the manipulation and visualization of point cloud data. Subjects were seated in a fixed position with point clouds positioned at a comfortable distance. They were allowed to rotate the point cloud clockwise and alternate between the reference and distorted point clouds. Empirically, it was decided that the evaluation required at least a full rotation and six commutations.

III. Results

The Mean Opinion Scores (MOS) obtained using the HMD (solid lines) and the 3D stereoscopic display [2] (dashed lines) are shown in fig. 2. The variations in the MOS with respect to the coding bitrate are very similar in both experiments. The plots in fig. 2 also show the 95% confidence interval (CI) of the subjective scores, considering a Gaussian distribution. The horizontal green line in each plot shows the resulting MOS for each reference-reference pair, and the corresponding CI is given by the green-shaded region. It should be noted that, in most cases, the MOS obtained using the HMD are lower than those obtained using a 3D display for the same content. This suggests that the subjects are more sensitive to artifacts created by the codecs, even at higher bitrates.

Fig. 3 shows scatter plots of paired MOS considering the evaluation using the HMD and each of the two previous evaluations, as well as the resulting linear fitting. Table III shows

[4]https://unity.com

[5]https://github.com/keijiro/Pcx

TABLE I: Point size for each content.

	V-PCC					G-PCC					RS-DLPCC					Draco				
Content	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05
Bumbameuboi	0.008					0.012	0.01				-	0.03	0.019	0.011	0.01	0.14	-	0.1	-	0.1
Guanyin	0.002					0.006	0.004	0.002			-	0.04	0.002			0.012	-	0.003	-	0.002
Longdress	0.002					0.007	0.003	0.002			-	0.04	0.002			0.013	-	0.004	-	0.002
Rhetorician	0.002					0.007	0.004	0.002			-	0.004	0.003	0.002		0.013	-	0.004	-	0.002
Romanoillamp	0.002					0.006	0.003	0.002			-	0.004	0.003	0.002		0.01	-	0.003	-	0.002
Soldier	0.002					0.007	0.004	0.003	0.002		-	0.004	0.002			0.012	-	0.003	-	0.002

TABLE II: Average of the 95% CIs for the three conducted experiments, considering a Gaussian distribution.

3D [2] vs HMD																					
		V-PCC					G-PCC					RS-DLPCC					Draco				
Test	**Global**	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05	-	R02	R03	R04	R05	R01	-	R03	-	R05
3D	**0.447**	0.441	0.404	0.417	0.428	0.411	0.549	0.509	0.436	0.431	0.407	-	0.571	0.466	0.466	0.418	0.420	-	0.417	-	0.403
HMD	**0.291**	0.397	0.375	0.348	0.428	0.232	0.304	0.359	0.436	0.235	0.120	-	0.371	0.363	0.357	0.300	0.249	-	0.294	-	0.044
2D [1] vs HMD (considering the first 16 subjects)																					
		V-PCC					G-PCC					RS-DLPCC					Draco				
Test	**Global**	R01	R02	R03	R04	R05	R01	R02	R03	R04	R05	-	R02	R03	R04	R05	R01	-	R03	-	R05
2D	**0.335**	0.396	0.421	0.378	0.339	0.401	0.307	0.409	0.357	0.355	0.294	-	0.245	0.393	0.369	0.350	0.094	-	0.351	-	0.241
HMD	**0.296**	0.418	0.378	0.310	0.339	0.241	0.320	0.346	0.357	0.246	0.128	-	0.389	0.366	0.371	0.315	0.261	-	0.290	-	0.048

TABLE III: Correlation statistics for the 2D [1] vs. HMD and 3D [2] vs. HMD comparisons.

Test	**PCC**	**SROCC**	**RMSE**	**OR**
2D VS HMD	0.943	0.942	0.109	0.539
3D VS HMD	0.934	0.924	0.118	0.588

TABLE IV: Kruskal-Wallis p-values [17] for the 2D [1] vs. HMD and 3D [2] vs. HMD comparisons.

Test	**Global**	**Draco**	**RS-DLPCC**	**G-PCC**	**V-PCC**
2D vs HMD	0.818	0.680	0.885	0.544	0.657
3D vs HMD	0.881	0.289	0.433	0.506	0.318

the corresponding performance indicators, notably the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Root-Mean Squared Error (RMSE), and the Outlier Ratio (OR) [18], revealing that the three tests result in a similar quality evaluation. Moreover, a Kruskal-Wallis one-way analysis followed by a multiple comparison test was performed [17] (table IV), revealing no statistical differences between the subjective evaluations ($p > 0.05$).

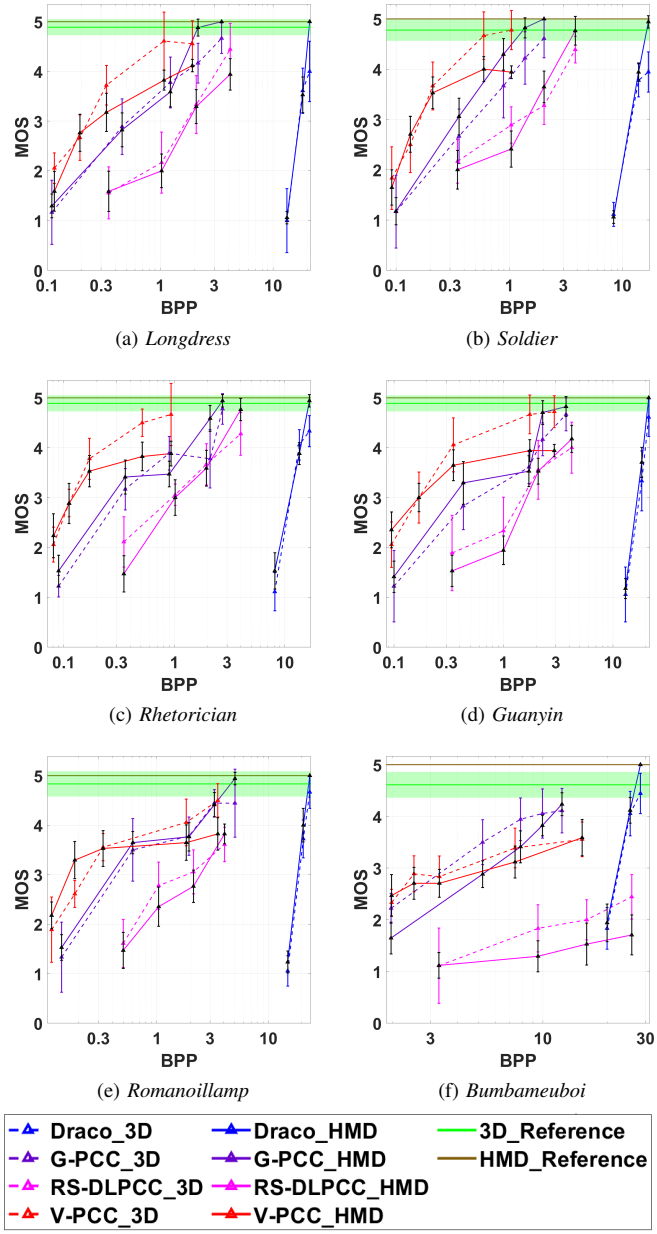
Table II shows the averages of the 95% CIs of the three quality evaluations taken across all source content for each codec-bitrate pair. In the *Global* column, the overall average of the 95% CIs is given. As the 2D evaluation only used 16 subjects, only the first 16 subjects were used to calculate the CIs to allow a direct comparison. All subjects were considered for the comparison with the 3D evaluation that used 18 subjects. The reported values indicate that the CIs are, on average, smaller in the evaluation using the HMD. This means that subjects tend to agree more on their scores for each stimulus. Furthermore, the scores tend to be lower for the higher rates, revealing that this methodology allows the differentiation on the high qualities. Finally, the hidden references were always given a score of 5 when using the HMD, which was not the case for the 2D and 3D displays.

TABLE V: Results of the multi-way ANOVA with repeated measures.

Source	**DF**	**F**	***p*-value**
Display	2	0.602	0.549
Content:Display	10	1.248	0.264
Compression Method:Display	6	6.018	< 0.001
Rate:Display	8	20.223	< 0.001

A multi-way ANOVA with repeated measures was also performed to further assess the impact of the display device on the obtained quality scores, although the data is not normally distributed. Nevertheless, it is relevant because it was shown to be robust to data that violates that assumption [19]. Furthermore, the impact of the interaction between the display type and the independent variables, i.e., content, codec (compression method), and distortion level (rate), was also assessed. The ANOVA considered only reference-distorted pairs and took the scores acquired for each pair with the different displays as repeated measures. Table V shows the summary of the multi-way ANOVA test. The p-value shows that the display does not have a statistically significant influence on the quality scores. The same result is observed when considering the interaction between display type and content. On the other hand, the interactions between the display type and both the codec and the rate yield p-values below 0.001, indicating that these factors have a statistically significant impact on the obtained scores.

It is noticeable from fig. 2 that the subjects were more sensitive to the artifacts using the HMD. Subjects revealed that they easily perceive distortions for higher bit rates, in opposition to the double stimulus tests used for comparison. Furthermore, the results from table II also show a greater CI's when evaluating different bit rates.

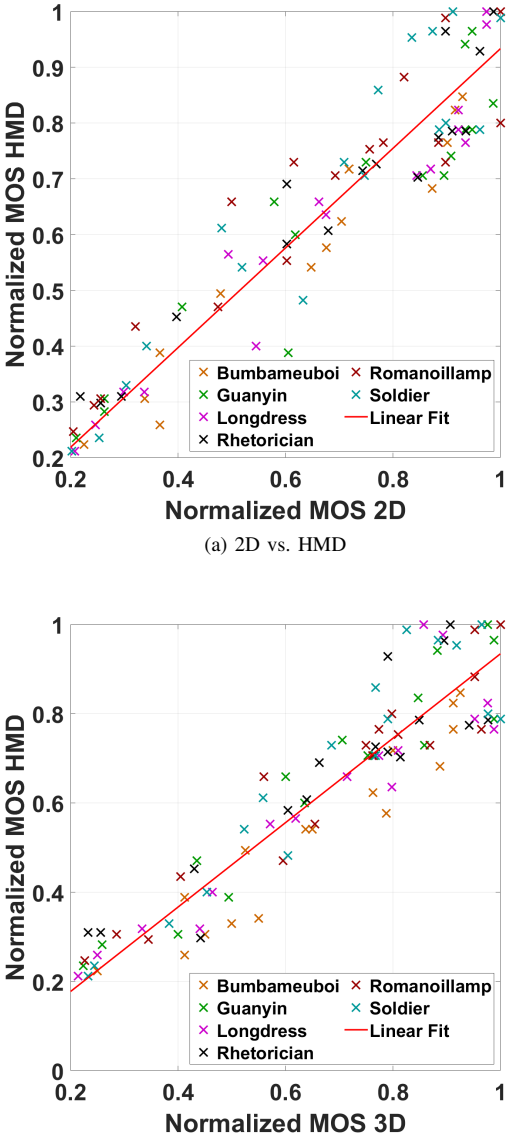


(a) *Longdress* (b) *Soldier*

(c) *Rhetorician* (d) *Guanyin*

(e) *Romanoillamp* (f) *Bumbameuboi*

Fig. 2: MOS vs. bpp with 95% CIs, considering a Gaussian distribution.



(a) 2D vs. HMD

(b) 3D vs. HMD

Fig. 3: Comparison between the MOS obtained using the 2D and the 3D representations, and respective linear fitting.

Finally, it is important to emphasize that different subjects were used in this study and in the preceding studies [1] and [2], preventing the influence of possible memory effects, as were observed in Testolina *et al.* [20].

IV. Conclusions

This paper reveals that the subjective evaluation model with a HMD yields highly correlated results with both 2D and 3D visualization. Furthermore, it offers several advantages, including narrower CIs, no need for computationally expensive video generation, with associated large uncompressed videos, and it requires less memory. This methodology, where users can switch between the distorted and reference content for a direct comparison, can also be considered for any displaying technology.

Although no time control was imposed to the subjects, the low confidence intervals seem to demonstrate that visualization time did not influence the quality judgment. Furthermore, requiring a complete rotation and switch six times between the reference and the distorted point clouds also revealed appropriate for the same reason.

REFERENCES

[1] J. Prazeres, M. Pereira, and A. Pinheiro, "Quality analysis of point cloud coding solutions," *Electronic Imaging*, 2022.

[2] J. Prazeres et al, "Subjective quality evaluation of point clouds with 3D stereoscopic visualization," in *ICIP*, 2022.

[3] E. Alexiou et al, "PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *QoMEX*, 2020.

[4] 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 *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020.

[5] X. Zhou et al, "QAVA-DPC: Eye-tracking based quality assessment and visual attention dataset for dynamic point cloud in 6 DoF," in *ISMAR*, 2023.

[6] X. Wu et al, "Subjective quality database and objective study of compressed point clouds with 6DoF head-mounted display," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[7] E. Alexiou et al, "Point cloud subjective evaluation methodology based on 2D rendering," in *QoMEX*, 2018.

[8] A. Evangelos et al, "Point cloud subjective evaluation methodology based on reconstructed surfaces," in *Applications of Digital Image Processing XLI*. SPIE, 2018.

[9] A. Javaheri et al, "Subjective and objective quality evaluation of compressed point clouds," in *MMSP*, 2017.

[10] L. A. da Silva Cruz et al, "Point cloud quality evaluation: Towards a definition for test conditions," in *QoMEX*, 2019.

[11] S. Perry et al, "Quality evaluation of static point clouds encoded using MPEG codecs," in *ICIP*, 2020.

[12] J. Prazeres et al, "Quality evaluation of machine learning-based point cloud coding solutions," in *Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis*, 2022.

[13] ——, "Quality evaluation of point cloud compression techniques," *Signal Processing: Image Communication*, 2024.

[14] D. Graziosi et al, "An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Transactions on Signal and Information Processing*, 2020.

[15] A. F. R. Guarda et al, "Deep learning-based point cloud geometry coding with resolution scalability," in *MMSP*, 2020.

[16] ITU-R BT.500-15, "Methodology for the subjective assessment of the quality of television pictures,," Jan 2012.

[17] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952. [Online]. Available: http://www.jstor.org/stable/2280779

[18] ITU-T P.1401, "International telecommunication union,," in *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*, Jul 2012.

[19] L. M. Lix et al, "Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance "f" test," *Review of Educational Research*, 1996.

[20] M. Testolina, D. Lazzarotto, R. Rodrigues, S. Mohammadi, J. Ascenso, A. M. Pinheiro, and T. Ebrahimi, "On the performance of subjective visual quality assessment protocols for nearly visually lossless image compression," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 6715–6723.

Bibliography

[1] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (v-pcc) and geometry-based (g-pcc)," *APSIPA Transactions on Signal and Information Processing*, vol. 9, p. e13, 2020. xxvii, 1, 3, 7, 17, 18, 19, 20, 21, 22, 28, 30, 60

[2] I. JTC1/SC29/WG1, "JPEG pleno point cloud use cases and requirements, v1.6, Doc. N100096," Jan 2022. xxvii, 3, 4, 10

[3] I. JTC1/SC29/WG11, "Use cases for point cloud compression (pcc), Doc. N16331," Jun 2016. xxvii, 6

[4] M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, *Autonomous driving: technical, legal and social aspects*. Springer Nature, 2016. xxvii, 7

[5] J. Prazeres, M. Pereira, and A. M. Pinheiro, "Quality evaluation of point cloud compression techniques," *Signal Processing: Image Communication*, vol. 128, p. 117156, 2024. xxvii, xxviii, xxix, 9, 13, 33, 47, 49, 53, 69, 72, 73, 76, 77

[6] J. Wang, D. Ding, Z. Li, and Z. Ma, "Multiscale point cloud geometry compression," in *2021 Data Compression Conference (DCC)*. IEEE, 2021, pp. 73–82. xxvii, 1, 23, 24, 57, 65

[7] M. Quach, G. Valenzise, and F. Dufaux, "Improved deep point cloud geometry compression," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, 2020, pp. 1–6. xxvii, 1, 23, 24, 31, 57, 60, 65

[8] A. F. Guarda, M. Ruivo, L. Coelho, A. Seleem, N. M. Rodrigues, and F. Pereira, "Deep learning-based point cloud coding and super-resolution: a joint geometry and color approach," *IEEE Transactions on Multimedia*, 2023. xxvii, 1, 23, 26, 73

[9] I. JTC1/SC29/WG1, "Verification model description for JPEG pleno learning-based point cloud coding v4.0, Doc. WG1N100709," Jan 2024. xxviii, 26, 27, 28, 29, 31

[10] J. Prazeres, Z. Luo, A. M. Pinheiro, L. A. da Silva Cruz, and S. Perry, "JPEG Pleno Call for Proposals responses quality assessment," in *ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2023, pp. 1–5. xxviii, 11, 12, 13, 33, 34, 65, 66, 67, 76, 77

[11] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE Computer Society, 2020, pp. 1–6. xxviii, 14, 35, 37, 64, 66, 71

[12] Y. Zhang, Q. Yang, and Y. Xu, "MS-GraphSIM: Inferring point cloud quality via multiscale graph similarity," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 1230–1238. xxviii, 39, 64, 66

[13] J. Prazeres, R. Rodrigues, M. Pereira, and A. M. Pinheiro, "Quality evaluation of machine learning-based point cloud coding solutions," in *Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis*, 2022, pp. 57–65. xxviii, 12, 13, 33, 55, 56, 65, 66, 67, 73, 76, 77

[14] J. Prazeres, M. Pereira, and A. M. Pinheiro, "Subjective quality evaluation of point clouds with 3D stereoscopic visualization," in *2022 IEEE international conference on image processing (ICIP)*. IEEE, 2022, pp. 2861–2865. xxviii, 12, 33, 54, 57, 75

[15] J. Prazeres, R. Rodrigues, M. Pereira, and A. M. Pinheiro, "Point cloud objective quality: Benchmarking features and quality evaluation," *arXiv preprint arXiv:2504.03381*, 2025. xxix, 59, 60, 67

[16] J. Prazeres, M. Pereira, and A. Pinheiro, "Quality analysis of point cloud coding solutions," *Electronic Imaging*, vol. 34, pp. 1–6, 2022. xxix, 12, 33, 54, 56, 65, 66, 67, 68, 69, 70, 71, 72, 73, 77

[17] I. JTC1/SC29/WG1, "JPEG pleno point cloud coding common training and test conditions v2.1, Doc. WG1N100841," Apr 2024. xxxi, 11, 52

[18] A. Ak, E. Zerman, M. Quach, A. Chetouani, A. Smolic, G. Valenzise, and P. Le Callet, "Basics: Broad quality assessment of static point clouds in a compression scenario," *IEEE Transactions on Multimedia*, vol. 26, pp. 6730–6742, 2024. xxxi, 33, 47, 60, 61, 69, 70

[19] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 3460–3464. xxxi, 13, 35, 36, 64, 66, 70, 73

[20] E. Alexiou and T. Ebrahimi, "Point cloud quality assessment metric based on angular similarity," in *2018 IEEE international conference on multimedia and expo (ICME)*. IEEE, 2018, pp. 1–6. xxxi, 70

[21] 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. xxxi, 42, 64, 66, 70

[22] E. Camuffo, D. Mari, and S. Milani, "Recent advancements in learning algorithms for point clouds: An updated overview," *Sensors*, vol. 22, no. 4, p. 1357, 2022. 2

[23] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg *et al.*, "The digital michelangelo project: 3d scanning of large statues," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 131–144. 2

[24] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building rome in a day," *Communications of the ACM*, vol. 54, no. 10, pp. 105–112, 2011. 2

[25] ITU-R BT.500-15, "Methodology for the subjective assessment of the quality of television pictures„" Jan 2023. 8, 32

[26] ITU-T P.1401, "Statistical analysis, evaluation and reporting guidelines of quality measurements," Jan 2020. 9, 47

[27] I. JTC1/SC29/WG1, "JPEG pleno point cloud coding common test conditions v3.6, Doc. WG1N91058," Apr 2021. 11

[28] ——, "Final call for evidence on JPEG pleno point cloud coding, Doc. WG1N88014," Jul 2020. 11

[29] ——, "Final call for proposals on JPEG pleno point cloud coding, Doc. WG1N100097," Jan 2022. 11, 26

[30] ——, "JPEG pleno point cloud scope and timeline v2, Doc. WG1 N100015," Oct 2021. 11, 23, 33

[31] J. Prazeres, R. Rodrigues, M. Pereira, and A. M. Pinheiro, "Subjective quality evaluation of point clouds using a head-mounted display," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5. 12, 75

[32] ——, "On the stability of point cloud machine learning based coding," in *2022 10th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2022, pp. 1–6. 12, 13, 72, 77

[33] M. 3DG, g-PCC Codec Description v5 ISO/IEC JTC1/SC29/WG11 N18891, Geneva, CH, October 2019. 13

[34] T. M. Borges, D. C. Garcia, and R. L. De Queiroz, "Fractional super-resolution of voxelized point clouds," *IEEE Transactions on Image Processing*, vol. 31, pp. 1380–1390, 2022. 13, 57, 65

[35] M. 3DG, v-PCC Codec Test Model v8 ISO/IEC JTC1/SC29/WG11 W18884, Geneva, CH, October 2019. 13

[36] J. Prazeres, R. Rodrigues, M. Pereira, and A. M. Pinheiro, "JPEG pleno learning-based point cloud coding: A performance analysis," in *2023 IEEE International*

Conference on Image Processing (ICIP). IEEE, 2023, pp. 1890–1894. 14, 56, 73, 77

[37] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, "Inferring point cloud quality via graph similarity," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 6, pp. 3015–3029, 2020. 14, 35, 39, 40, 64, 66, 71

[38] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *ITU - Telecommunications Standardization Sector: VCEG-M33*, March 2001. 14

[39] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15, no. 4, pp. 353–363, 1993. 15, 16

[40] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004. 15

[41] J. Geng, "Structured-light 3D surface imaging: a tutorial," *Advances in optics and photonics*, vol. 3, no. 2, pp. 128–160, 2011. 16

[42] C. Mallet and F. Bretar, "Full-waveform topographic lidar: State-of-the-art," *ISPRS Journal of photogrammetry and remote sensing*, vol. 64, no. 1, pp. 1–16, 2009. 16

[43] N. Haala, M. Kölle, M. Cramer, D. Laupheimer, and F. Zimmermann, "Hybrid georeferencing of images and lidar data for uav-based point cloud collection at millimetre accuracy," *ISPRS Open Journal of Photogrammetry and Remote Sensing*, vol. 4, p. 100014, 2022. 16

[44] V. Roback, A. Bulyshev, F. Amzajerdian, and R. Reisse, "Helicopter flight test of 3d imaging flash lidar technology for safe, autonomous, and precise planetary landing," in *Laser Radar Technology and Applications XVIII*, vol. 8731. SPIE, 2013, pp. 129–148. 16

[45] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3d," in *ACM siggraph 2006 papers*, 2006, pp. 835–846. 17

[46] "Lytro illum 40 megaray light field camera." [Online]. Available: https://www.dpreview.com/products/lytro/compacts/lytro_illum 17

[47] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926–954, 2017. 17

[48] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (pcl)," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 1–4. 17

[49] R. L. De Queiroz and P. A. Chou, "Compression of 3D point clouds using a region-adaptive hierarchical transform," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3947–3956, 2016. 18, 60

[50] J. Ascenso, P. Akyazi, F. Pereira, and T. Ebrahimi, "Learning-based image coding: early solutions reviewing and subjective quality evaluation," in *Optics, Photonics and Digital Technologies for Imaging Applications VI*, vol. 11353. SPIE, 2020, pp. 164–176. 23, 31

[51] A. F. Guarda, N. M. Rodrigues, and F. Pereira, "Point cloud coding: Adopting a deep learning-based approach," in *2019 Picture Coding Symposium (PCS)*. IEEE, 2019, pp. 1–5. 23

[52] ——, "Deep learning-based point cloud geometry coding with resolution scalability," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2020, pp. 1–6. 23, 25, 33, 65

[53] ——, "Adaptive deep learning-based point cloud geometry coding," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 2, pp. 415–430, 2020. 23, 25, 57, 65

[54] J. Wang, D. Ding, Z. Li, X. Feng, C. Cao, and Z. Ma, "Sparse tensor-based multiscale representation for point cloud geometry compression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 9055–9071, 2022. 23

[55] G. Valenzise, M. Quach, D. Tian, J. Pang, and F. Dufaux, "Point cloud compression," in *Immersive Video Technologies*. Elsevier, 2023, pp. 357–385. 23

[56] C. Zhang, D. Florencio, and C. Loop, "Point cloud attribute compression with graph transform," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 2066–2070. 23

[57] J. Zhang, J. Wang, D. Ding, and Z. Ma, "Scalable point cloud attribute compression," *IEEE Transactions on Multimedia*, 2023. 23, 26

[58] S. Gu, J. Hou, H. Zeng, H. Yuan, and K.-K. Ma, "3D point cloud attribute compression using geometry-guided sparse representation," *IEEE Transactions on Image Processing*, vol. 29, pp. 796–808, 2019. 23

[59] S. Gu, J. Hou, H. Zeng, and H. Yuan, "3D point cloud attribute compression via graph prediction," *IEEE Signal Processing Letters*, vol. 27, pp. 176–180, 2020. 23

[60] D. T. Nguyen and A. Kaup, "Lossless point cloud geometry and attribute compression using a learned conditional probability model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 8, pp. 4337–4348, 2023. 23

[61] J. Wang and Z. Ma, "Sparse tensor-based point cloud attribute compression," in *2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2022, pp. 59–64. 23

[62] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017. 23, 27

[63] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su *et al.*, "Shapenet: An information-rich 3D model repository," *arXiv preprint arXiv:1512.03012*, 2015. 23

[64] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D shapenets: A deep representation for volumetric shapes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1912–1920. 24

[65] M. Quach, G. Valenzise, and F. Dufaux, "Folding-based compression of point cloud attributes," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 3309–3313. 25

[66] ITU-T H., "High efficiency video coding," in *International Telecommunication Union*, Nov 2019. 25

[67] E. Alexiou, K. Tung, and T. Ebrahimi, "Towards neural network approaches for point cloud compression," in *Applications of digital image processing XLIII*, vol. 11510. SPIE, 2020, pp. 18–37. 25

[68] X. Sheng, L. Li, D. Liu, Z. Xiong, Z. Li, and F. Wu, "Deep-PCAC: An end-to-end deep lossy compression framework for point cloud attributes," *IEEE Transactions on Multimedia*, vol. 24, pp. 2617–2632, 2021. 25

[69] Z. Guo, Y. Zhang, L. Zhu, H. Wang, and G. Jiang, "TSC-PCAC: Voxel transformer and sparse convolution-based point cloud attribute compression for 3d broadcasting," *IEEE Transactions on Broadcasting*, 2024. 25

[70] G. Fang, Q. Hu, H. Wang, Y. Xu, and Y. Guo, "3DAC: Learning attribute compression for point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14 819–14 828. 26

[71] C. Choy, J. Gwak, and S. Savarese, "4D spatio-temporal convnets: Minkowski convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3075–3084. 27, 29

[72] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," *Advances in neural information processing systems*, vol. 31, 2018. 28

[73] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432. 29

[74] I. JTC1/SC29/WG1, "JPEG AI verification model software guidelines,Doc. WG1N100780," Jan 2024. 31

[75] E. Alexiou, T. Ebrahimi, M. V. Bernardo, M. Pereira, A. Pinheiro, L. A. D. S. Cruz, C. Duarte, L. G. Dmitrovic, E. Dumic, D. Matkovics *et al.*, "Point cloud subjective evaluation methodology based on 2D rendering," in *2018 Tenth international conference on quality of multimedia experience (QoMEX)*. IEEE, 2018, pp. 1–6. 33, 54

[76] E. Alexious, A. M. Pinheiro, C. Duarte, D. Matković, E. Dumić, L. A. da Silva Cruz, L. G. Dmitrović, M. V. Bernardo, M. Pereira, and T. Ebrahimi, "Point cloud subjective evaluation methodology based on reconstructed surfaces," in *Applications of Digital Image Processing XLI*, vol. 10752. SPIE, 2018, pp. 160–173. 33

[77] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of 3D point cloud denoising algorithms," in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2017, pp. 1–6. 33

[78] L. A. da Silva Cruz, E. Dumić, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation: Towards a definition for test conditions," in *2019 Eleventh international conference on quality of multimedia experience (QoMEX)*. IEEE, 2019, pp. 1–6. 33, 54

[79] H. Su, Z. Duanmu, W. Liu, Q. Liu, and Z. Wang, "Perceptual quality assessment of 3D point clouds," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3182–3186. 33

[80] E. Alexiou, I. Viola, T. M. Borges, T. A. Fonseca, R. L. De Queiroz, and T. Ebrahimi, "A comprehensive study of the rate-distortion performance in MPEG point cloud compression," *APSIPA Transactions on Signal and Information Processing*, vol. 8, p. e27, 2019. 33, 42

[81] 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 *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 3428–3432. 33, 54, 73

[82] E. Alexiou, E. Upenik, and T. Ebrahimi, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *2017 IEEE 19th international workshop on multimedia signal processing (MMSP)*. IEEE, 2017, pp. 1–6. 33

[83] 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 *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 127–136. 33

[84] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828–842, 2016. 33

[85] E. Alexiou, N. Yang, and T. Ebrahimi, "PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6. 33, 57

[86] X. Zhou, I. Viola, E. Alexiou, J. Jansen, and P. Cesar, "QAVA-DPC: Eye-tracking based quality assessment and visual attention dataset for dynamic point cloud in 6 DoF," in *2023 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE, 2023, pp. 69–78. 33

[87] X. Wu et al, "Subjective quality database and objective study of compressed point clouds with 6DoF head-mounted display," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021. 33

[88] S. Perry, L. A. Da Silva Cruz, J. Prazeres, A. Pinheiro, E. Dumic, D. Lazzarotto, and T. Ebrahimi, "Subjective and objective testing in support of the jpeg pleno point cloud compression activity," in *2022 10th European Workshop on Visual Information Processing (EUVIP)*, 2022, pp. 1–6. 33

[89] Q. Liu, H. Su, Z. Duanmu, W. Liu, and Z. Wang, "Perceptual quality assessment of colored 3D point clouds," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 8, pp. 3642–3655, 2022. 33, 65, 66

[90] 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*, vol. 23, pp. 3877–3891, 2020. 33, 65, 66

[91] Y. Liu, Q. Yang, Y. Xu, and L. Yang, "Point cloud quality assessment: Dataset construction and learning-based no-reference metric," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 2s, pp. 1–26, 2023. 33, 44

[92] L. Hua, M. Yu, Z. He, R. Tu, and G. Jiang, "CPC-GSCT: Visual quality assessment for coloured point cloud based on geometric segmentation and colour transformation," *IET image processing*, vol. 16, no. 4, pp. 1083–1095, 2022. 33

[93] S. Perry, L. A. D. S. Cruz, E. Dumic, N. H. T. Nguyen, A. Pinheiro, and E. Alexiou, "Comparison of remote subjective assessment strategies in the context of the jpeg pleno point cloud activity," in *2021 IEEE 23rd international workshop on Multimedia signal processing (MMSP)*. IEEE, 2021, pp. 1–6. 33

[94] D. Lazzarotto, M. Testolina, and T. Ebrahimi, "Subjective performance evaluation of bitrate allocation strategies for MPEG and JPEG Pleno point cloud compression," *EURASIP Journal on Image and Video Processing*, vol. 2024, no. 1, p. 14, 2024. 33, 47

[95] 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*, vol. 10752. SPIE, 2018, pp. 174–190. 36, 40, 64, 66

[96] BT709 ITU-R BT.70, "Parameter values for the HDTV standards for production and international programme exchange," Jun 2015. 36

[97] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in *ICMEW*, 2020. 38, 64, 66

[98] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "A point-to-distribution joint geometry and color metric for point cloud quality assessment," in *2021 IEEE 23rd international workshop on Multimedia signal processing (MMSP)*. IEEE, 2021, pp. 1–6. 41, 64, 66

[99] ——, "Mahalanobis based point to distribution metric for point cloud geometry quality evaluation," *IEEE Signal Processing Letters*, vol. 27, pp. 1350–1354, 2020. 41

[100] I. Viola, S. Subramanyam, and P. Cesar, "A color-based objective quality metric for point cloud contents," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6. 41, 64

[101] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 1, pp. 194–201, 2011. 42

[102] Q. Yang, Y. Liu, S. Chen, Y. Xu, and J. Sun, "No-reference point cloud quality assessment via domain adaptation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 21 179–21 188. 45

[103] A. D. B. C. N. Network, "Blind image quality assessment using a deep bilinear convolutional neural network," *Deep Bilinear Convolutional Neural*, 2022. 45

[104] Z. Zhang, W. Sun, X. Min, Q. Wang, J. He, Q. Zhou, and G. Zhai, "MM-PCQA: Multimodal learning for no-reference point cloud quality assessment," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 2023, pp. 1759–1767. 45, 46

[105] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, and G. Zhai, "No-reference quality assessment for 3D colored point cloud and mesh models," *IEEE Transactions on*

Circuits and Systems for Video Technology, vol. 32, no. 11, pp. 7618–7631, 2022. 46

[106] M. Awad and R. Khanna, "Support vector regression," in *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. Springer, 2015, pp. 67–80. 46, 59, 60

[107] Z. Zhang, W. Sun, Y. Zhu, X. Min, W. Wu, Y. Chen, and G. Zhai, "Evaluating point cloud from moving camera videos: A no-reference metric," *IEEE Transactions on Multimedia*, 2023. 46

[108] 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, 2004. 47

[109] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The thrity-seventh asilomar conference on signals, systems & computers, 2003*, vol. 2. IEEE, 2003, pp. 1398–1402. 47

[110] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006. 47

[111] 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, 2011. 47

[112] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," *The Netflix Tech Blog*, vol. 6, no. 2, p. 2, 2016. 47

[113] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595. 47

[114] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. G. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for HDR image quality assessment," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, p. 39, 2015. 47, 67

[115] D. Lazzarotto, E. Alexiou, and T. Ebrahimi, "Benchmarking of objective quality metrics for point cloud compression," in *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2021, pp. 1–6. 47, 67

[116] K. Pearson, "Vii. note on regression and inheritance in the case of two parents," *proceedings of the royal society of London*, vol. 58, no. 347-352, pp. 240–242, 1895. 47

[117] ——, "Mathematical contributions to the theory of evolution. viii. on the correlation of characters not quantitatively measurable," *Proceedings of the Royal Society of London*, vol. 66, no. 424-433, pp. 241–244, 1900. 47

[118] C. Spearman, "The proof and measurement of association between two things." 1961. 47

[119] J. Kenney, *Characteristic functions in statistics*. JSTOR, 1942, vol. 17, no. 2. 48

[120] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John wiley & sons, 2003. 48

[121] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1-2, pp. 81–93, 1938. 48

[122] Z. Taylor, J. Nieto, and D. Johnson, "Multi-modal sensor calibration using a gradient orientation measure," *Journal of Field Robotics*, vol. 32, no. 5, pp. 675–695, 2015. 51

[123] I. H.264, "Recommendation h.264," Aug 2021. 55

[124] J. Prazeres, R. Rodrigues, M. Pereira, and A. M. Pinheiro, "Performance analysis of deep learning-based lossy point cloud geometry compression coding solutions," *IEEE Access*, 2025. 57, 73, 76, 77

[125] "ISO/IEC JTC1/SC29/WG1M100119, FSM: quality assessment of point clouds based on quality fea- tures selection." 59

[126] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970. 59, 60

[127] H. Drucker, "Improving regressors using boosting techniques," in *Icml*, vol. 97, no. 107, 1997, p. e115. 59

[128] O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Overfitting, model tuning, and evaluation of prediction performance*. Springer, 2022, pp. 109–139. 63

[129] W. Zhou, G. Yue, R. Zhang, Y. Qin, and H. Liu, "Reduced-reference quality assessment of point clouds via content-oriented saliency projection," *IEEE Signal Processing Letters*, vol. 30, pp. 354–358, 2023. 64, 66

[130] R. Watanabe, S. N. Sridhara, H. Hong, E. Pavez, K. Nonaka, T. Kobayashi, and A. Ortega, "Full reference point cloud quality assessment using support vector regression," *Signal Processing: Image Communication*, vol. 131, p. 117239, 2025. 64, 66

[131] ITU-T P.1401, "International telecommunication union„" in *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*, Jul 2012. 67

[132] P. Bo, R. Ling, and W. Wang, "A revisit to fitting parametric surfaces to point clouds," *Computers & Graphics*, vol. 36, no. 5, pp. 534–540, 2012, shape Modeling International (SMI) Conference 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0097849312000751 69

[133] "ISO/IEC JTC1/SC29/WG1M89044, JPEG Pleno PC exploration study 4 results." 69

[134] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American statistical Association*, vol. 47, no. 260, pp. 583–621, 1952. 69

[135] L. M. Lix, J. C. Keselman, and H. J. Keselman, "Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance f test," *Review of educational research*, vol. 66, no. 4, pp. 579–619, 1996. 69, 70

[136] L. Krasula, K. Fliegel, P. Le Callet, and M. Klíma, "On the accuracy of objective image and video quality models: New methodology for performance evaluation," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2016, pp. 1–6. 69

[137] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The annals of mathematical statistics*, pp. 50–60, 1947. 70

[138] J. W. Tukey, "Comparing individual means in the analysis of variance," *Biometrics*, pp. 99–114, 1949. 71

[139] J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases." *Radiology*, vol. 148, no. 3, pp. 839–843, 1983. 71

[140] R. A. Fisher, "On the interpretation of χ 2 from contingency tables, and the calculation of p," *Journal of the royal statistical society*, vol. 85, no. 1, pp. 87–94, 1922. 71