QoE Modeling in Volumetric Video Streaming: A Short Survey

Mojtaba Mozhganfar*, Masoumeh Khodarahmi†, Daniele Lorenzi‡,
Mahdi Dolati§, Farzad Tashtarian‡, Ahmad Khonsari*, Christian Timmerer‡

*School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran
†IMDEA Networks Institute, Leganes, Spain
‡Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität Klagenfurt, Austria
§Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

***Abstract*—Volumetric video streaming enables six degrees of freedom (6DoF) interaction, allowing users to navigate freely within immersive three-dimensional (3D) environments. Despite notable advancements, volumetric video remains an emerging field, presenting ongoing challenges and vast opportunities in content capture, compression, transmission, decompression, rendering, and display. As user expectations grow, delivering high Quality of Experience (QoE) in these systems becomes increasingly critical due to the complexity of volumetric content and the demands of interactive streaming. This paper reviews recent progress in QoE for volumetric streaming, beginning with an overview of QoE evaluation of volumetric video streaming studies, including subjective assessments tailored to 6DoF content. The core focus of this work is on objective QoE modeling, where we analyze existing models based on their input factors and methodological strategies. Finally, we discuss the key challenges and promising research directions for building perceptually accurate and adaptable QoE models that can support the future of immersive volumetric media.**

***Index Terms*—Volumetric video streaming, QoE, QoE modeling, subjective assessment, objective metrics, immersive media.**

I. Introduction

Volumetric video enables immersive and interactive experiences by capturing real-world scenes in three-dimensional (3D) formats, such as point clouds, meshes, or multi-view videos with depth. By delivering large volumes of 3D data in real-time, volumetric video streaming (VVS) supports features such as free-viewpoint navigation and six degrees of freedom (6DoF) interaction [1]. These capabilities make VVS suitable for a wide range of applications, including augmented reality (AR) and virtual reality (VR), education, entertainment, and healthcare [2]. As interest in these domains continues to grow, volumetric content is receiving increasing attention from both industry and academia. This trend is also reflected in market projections, which estimate that the volumetric video sector will reach USD 24.23 billion by 2032 [3].

Despite its potential, delivering volumetric video presents significant technical challenges. VVS systems are highly resource-intensive because they require high bandwidth, low latency, and substantial processing power on both servers and clients. In addition, they must account for user-driven interactivity and unpredictable access patterns [4,5]. Because network behavior is highly variable and hard to model in full detail, many VVS works use simplified delivery-side abstractions, such as measured throughput, simple end-to-end delay, and stall statistics, and rely on these as core inputs for Quality of Experience (QoE) modeling and adaptation decisions [4]–[7]. These challenges make conventional streaming methods inadequate and highlight the need to rethink the design, implementation, and evaluation of immersive media systems [8]–[10].

Addressing these challenges requires a clear understanding of QoE to balance user satisfaction with scalability, for instance, by managing resolution and frame-rate trade-offs [11]. In practice, QoE models are often embedded directly into the streaming control loop, where they act as objectives and constraints for viewport-driven bitrate and tile selection, rate allocation, and adaptation decisions [6,7]. Yet QoE modeling in VVS remains difficult because many two-dimensional (2D) metrics, such as the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM), do not capture 6DoF-specific perceptual and interactive aspects such as viewpoint changes, temporal stability, interaction latency, and realism [12]–[15]. We therefore provide a short survey of QoE modeling in VVS, summarize key influencing factors (IFs) and modeling strategies, and outline challenges and directions for future work.

II. Background

A. Volumetric Video Streaming (VVS)

The typical VVS pipeline consists of six main stages: content capture, compression, transmission, decompression, rendering, and display [10,16]. VVS content uses various 3D formats, each with trade-offs in compression, rendering, and streaming. Common representations include point clouds [17], plenoptic point clouds [18], polygonal meshes [19], voxels [20], implicit surfaces [21], neural radiance fields (NeRFs) [22], and 3D Gaussian Splatting (3DGS) [23]. High-resolution scenes are captured with panoramic sensor arrays and compressed using video coding standards developed by Moving Picture Experts Group (MPEG) for multimedia applications [24], notably MPEG Point Cloud Compression (MPEG-PCC) for volumetric content [25].

Clients decode and render the content in real-time, typically for head-mounted displays. To support immersive 6DoF interaction, *e.g.*, adjusting viewpoint (x, y, z) and orientation $(\texttt{yaw}, \texttt{pitch}, \texttt{roll})$, VVS systems must maintain low-latency operation throughout the pipeline. This is particularly challenging due to the large data volume of volumetric representations and the tight interactive deadlines, which amplify bandwidth and client-side compute demands [16,26,27]. Accordingly, VVS systems often rely on adaptation and optimization across the pipeline based on network variability and device capabilities [8,9].

B. QoE

QoE measures the end-user's perceived satisfaction with a service, influenced by technical performance (*e.g.*, network, media quality), contextual factors, and subjective expectations [28]. According to ITU-T Recommendation P.1201, a QoE model is defined as "an algorithm with the purpose of estimating the subjective (perceived) quality of a media sequence" [29]. Early QoE models, which aim to accurately predict the perceived quality from the end-user's perspective by incorporating a range of IFs, can be categorized into different types such as bitstream, parametric, and hybrid models based on the nature of the input data [30,31]. QoE IFs are measurable aspects of the user, system, service, application, or context that can impact the perceived QoE [32]. The IFs are broadly categorized into human-related, system-related, and context-related classes [32], with an additional content-related class introduced for video applications [33]. Various models have been proposed using different inputs and approaches to estimate the impact of quality variations and stalling events [30,34].

QoE modeling in 2D video streaming has been extensively studied [35,36], particularly in the context of HTTP Adaptive Streaming (HAS) technologies such as MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) and HTTP Live Streaming (HLS) [30]. Various techniques to model QoE exist, including data-driven approaches [37]–[39] and machine learning (ML) frameworks [40]–[42]. Among the IFs, stalling events and long startup delays are consistently reported as major causes of dissatisfaction [43]. With the increasing adoption of immersive media technologies, such as AR/VR and telepresence, there is a growing need to extend QoE modeling approaches to support volumetric and 3D video streaming, where QoE models for VVS differ from traditional 2D-centric models [14].

III. QoE in VVS

In VVS, users are highly sensitive to visual quality, latency, and system responsiveness due to the immersive and interactive nature of 3D content. Consequently, QoE is central to the success and adoption of VVS. Research in this area typically falls into two categories: subjective studies, which capture user perception but are hard to scale, and objective models, which estimate QoE from measurable system and content features [59]. While objective methods are more practical, they must be validated against subjective assessments, which provide the ground truth. The following subsections review key studies on QoE in VVS.

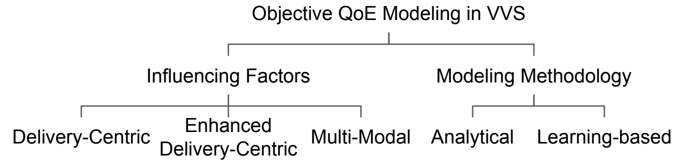

Fig. 1. Objective QoE Modeling in VVS

A. Subjective QoE Assessment for VVS

Subjective assessment captures user-perceived quality, usually via mean opinion scores (MOS) [60]. Several datasets have been developed based on subjective QoE assessment processes, where user studies were conducted to collect quality ratings of volumetric video content under different conditions [15,48,50,53,55,61]. For instance, Wu *et al.* [61] conducted a large study in which participants explore VR point cloud scenes and rate their quality. These datasets often include MOS scores from human participants who view rendered sequences subjected to various compression and transmission conditions. Although subjective tests provide direct insights, they are costly and hard to scale [62]. Therefore, many QoE models use objective methods trained on subjective data, estimating QoE from different factors to balance accuracy with practicality. Consequently, subjective results are used for training and validating QoE models.

B. Objective QoE Modeling in VVS

Objective QoE modeling can be investigated from two distinct perspectives: (i) the manner in which IFs are considered, (ii) the methodology used to model QoE itself as it shown in Figure 1. We also provide a summary of existing QoE modeling studies for VVS, as presented in Table I.

1) IF: Among VVS representations, point clouds are used most often due to their simpler structure, lower processing cost, and stronger support for real-time transmission [16,63]. This trend is reinforced by standardization efforts such as MPEG's video-based PCC (V-PCC) [64] and geometry-based PCC (G-PCC) [65], which provide efficient encoding and delivery for point cloud data. While a few studies model QoE for mesh-based content [7], most work still targets point clouds because they are easier to deploy and more common in practice. Traditional QoE model categories in HAS, such as parametric, bitstream, and hybrid models [30], were designed for 2D video. They can still be used for volumetric content, but they need adjustments to capture 3D-specific characteristics. Next, we review QoE models for point clouds and classify them into three groups:

Delivery-centric models. Inspired by in-service QoE monitoring for IP-based audiovisual services [31,66] and prior QoE taxonomies for adaptive streaming [30], we define delivery-centric (DC) models as those that estimate QoE only from delivery and runtime signals, without parsing the volumetric bitstream and without reconstructing the media content. In other words, all inputs are obtainable from packet headers,

TABLE I

COMPARING VVS QOE MODELS OVER DIFFERENT INFLUENCING FACTORS, MODELING BASED ON IF TYPES (DELIVERY-CENTRIC (DC), ENHANCED DELIVERY-CENTRIC (EDC), MULTI-MODAL (MM)), MODELING METHODOLOGIES (ANALYTICAL (A), LEARNING-BASED (LB)), APPLICATION SCENARIOS (LIVE, VIDEO-ON-DEMAND (VOD)), AND THE AVAILABILITY OF PUBLIC DATASETS.

Study	**Year**	**IF set**	**Modeling**		**Type**	**Dataset**
			IF type	**Method**		
A. Zhang *et al.* [44]	2021	visual quality, stall duration, quality variation	MM	A	VoD	✗
J. Li *et al.* [15]	2022	PCPSNR, stall, quality switch	MM	A	VoD	[45]
A. Zhang *et al.* [46]	2022	point density, viewing distance, SR ratio and distortion, stall duration, quality switch	MM	A	VoD	✗
Y. Liu *et al.* [4]	2022	viewport (drift, smoothness, movement distance), latency, quality variation, stall duration	DC	LB	VoD	✗
X. Wang *et al.* [47]	2023	V-PSNR, stall, quality variation	MM	A	VoD	✗
J. Li *et al.* [6]	2023	PSNR, stall duration, stall frequency	MM	A	VoD	✗
J. Weil *et al.* [48]	2023	frame rate, QP, objects distance	EDC	LB	VoD	[49]
J. Li *et al.* [50]	2023	PSNR, stall, quality switch	MM	A	Live	[51]
Y. Gao *et al.* [52]	2023	PSNR, encoding level, objects distance, stall, quality smoothness, decoding time	MM	A	VoD	✗
M. Nguyen *et al.* [53]	2023	geometry and texture QP, quality switch, viewing distance	EDC	LB	VoD	[54]
D. V. Nguyen *et al.* [55]	2024	geometry and texture QP, chunks bitrate	EDC	A	VoD	[56]
D. V. Nguyen *et al.* [55]	2024	chunks bitrate, stall	DC	LB	VoD	[56]
Y. Shi *et al.* [57]	2024	geometry and texture QP	EDC	LB	VoD	✗
Y. Shi *et al.* [11]	2024	geometry and texture QP, viewpoint of user	EDC	LB	VoD	✗
B. Li *et al.* [7]	2024	visual quality, quality variation (fluctuation or smoothness), playback delay (latency)	MM	A	Live	✗
K. Hu *et al.* [58]	2025	visual quality, quality variation, startup delay	MM	A	Live	✗
C. Wang *et al.* [9]	2025	visual quality, quality variation, stall	MM	A	VoD	✗

transport statistics, and player/system logs, such as delay and loss, throughput and buffer level, startup delay, stall events, and quality switches, as well as interaction traces when available [66,67]. We assign a paper to DC if its reported QoE inputs are limited to these delivery-side signals. In VVS, these QoE models leverage factors like quality switch and stall [4,55,68], latency [4], and startup delay [59] for adaptive delivery [58].

Enhanced Delivery-centric models. Building on bitstream-based QoE assessment concepts [30,31,66], we define enhanced delivery-centric (EDC) models as DC models augmented with features parsed from the encoded volumetric stream, while still avoiding access to the original reference and full content reconstruction. Concretely, EDC models extract codec and encoder descriptors from volumetric bitstreams (*e.g.*, V-PCC, G-PCC), such as bitrate, geometry/texture quantization parameters (QP), and frame rate [16,48,57]. We assign a paper to EDC if it uses at least one bitstream-derived feature, even when it also uses delivery-side logs. For example, Nguyen *et al.* [55] learn a weighted QoE predictor from bitstream features via least squares, and Shi *et al.* [57] separate geometry and texture compression effects to explain perceived quality changes.

Multi-modal models. Following the concept introduced in [30,31,66], we define multi-modal approaches as methods that combine content features and network conditions to support more comprehensive and adaptive QoE prediction. These models incorporate various types of information such as visual quality metrics, packet-level data, bitstream features, and reconstructed media content to capture the joint effects of encoding settings, network impairments, and content complexity. In VVS, multi-modal QoE models typically integrate factors such as visual quality, stall duration, quality variation, startup delay, and quality switching [47,50,58]. For example, Zhang *et al.* [44] and Wang *et al.* [9] propose models for video-on-demand (VoD) scenarios that use objective video quality metrics together with stall duration and quality variation to estimate QoE. Similarly, Hu *et al.* [58] focus on live streaming and emphasize the role of startup delay and quality variation in QoE assessment. Li *et al.* [15] introduce a dedicated QoE model for point cloud streaming that reflects the flexibility of multi-modal approaches. Their model incorporates point cloud PSNR (PCPSNR), a point cloud-specific visual quality metric, alongside other key factors such as stall duration and quality switching. By integrating these diverse inputs, their model adapts well to the dynamic conditions of VVS and achieves improved prediction accuracy.

2) Methodology: On the other hand, models in VVS are primarily analytical [15] or learning-based [53,59], depending on their structure. Some models from the input-based taxonomy are revisited here by their modeling methods, so certain works appear in both classifications to reflect their multi-dimensional contributions.

Analytical models. These models define QoE using factors such as visual quality, stall duration, and latency, and analytically formulate it as a function of these parameters to enable optimization through mathematical methods. Many of them apply heuristics after formulating the QoE to improve performance under constraints. For example, Zhang *et al.* [44] select super resolution (SR) levels by combining visible patch quality with spatial and temporal penalties. Hu *et al.* [58] propose chunk-level bitrate decisions are guided by control heuristics, while X. Wang *et al.* [47] use perceptual weighting to prioritize tile selection. Subramanyam *et al.* [61] explore heuristic bitrate allocation based on user viewport. Some analytical models adopt formal optimization techniques for QoE enhancement. L. Wang *et al.* [68] apply a greedy algorithm for near-optimal performance with low complexity. Li *et al.* [7] uses dynamic programming for rate adaptation, while C. Wang *et al.* [9] employ model predictive control

(MPC) for quality management.

Overall, analytical models formulate QoE by explicitly capturing the effects of spatial and temporal variation, playback stalls, and key perceptual factors under limited resources. Some studies separate *pure* analytical formulations solved by classical optimization using a solver such as nonlinear programming in [6] from *hybrid* methods that keep an explicit analytical QoE objective but rely on learning-based decision making, for instance rolling optimization solved by a DRL agent in [50]. Due to space limits, we treat both pure and hybrid methods as *analytical* models in this survey.

Learning-based models. These models use ML trained on subjective QoE data to predict perceived quality or optimize streaming. Shi *et al.* [57] use polynomial regression based on compression levels, building on the VOLVQAD dataset [69] and extended in QV4 [11] with viewpoint-aware tiling and predictions using gated recurrent units (GRUs). Similarly, Nguyen *et al.* [55] propose least squares regression and gradient tree boosting (GTB) models to assess multiple QoE factors. Other studies use learning-based methods such as ensembles, reinforcement learning (RL), and linear models trained on subjective data. Both [48] and [53] compare five models and report that GTB performs best, despite using different features. The former uses frame rate, compression level, and viewing distance, while the latter uses six inputs: start and end geometry QP, start and end texture QP, viewing distance, and bitrate. Liu *et al.* [4] analyze key volumetric QoE factors and compare four supervised learning models trained on user data. Simple regression models, such as [15], estimate QoE by weighting quality changes based on the tile's distance from the user. Many learning-based studies primarily report prediction accuracy; however, they do not explicitly quantify inference or feature-extraction cost. Consequently, ignoring compute can be risky, as compute-agnostic designs can degrade QoE. Learning-based models can capture complex patterns and adapt to diverse scenarios with sufficient data, while analytical models are simpler and more efficient.

IV. Opportunities and Challenges

QoE modeling in VVS offers clear opportunities (*e.g.*, multimodal integration, temporal modeling), yet it remains constrained by limited factor coverage, computational overhead, and weak generalization across volumetric formats. We summarize key open challenges that can guide future work toward more complete and deployment-oriented QoE models.

I) Limited Generalizability Across Volumetric Formats: Most QoE models focus on point clouds [17], which are also the most widely used volumetric representation in practice [70, 71]. In contrast, QoE modeling remains limited for other formats, including plenoptic point clouds [18], meshes [19], voxels [20], implicit surfaces [21], NeRF [22], and 3DGS [23], which reduces the applicability of existing assessments.

II) Temporal Dynamics Modeling: LSTM-based models capture temporal dependencies effectively in HAS [72], but comparable temporal modeling is still largely unexplored for volumetric QoE. Given time-varying 6DoF interactions and streaming variability, incorporating temporal dynamics can improve prediction of QoE evolution.

III) Real-Time, Scalable QoE Prediction and Standardization: Lightweight, real-time QoE models with low computational cost are essential for deployment on edge and mobile devices. In parallel, standardized subjective test protocols, including test environments, viewer instructions, and questionnaires, would strengthen validation and enable more consistent benchmarking of volumetric QoE models.

IV) Limited and Fragmented Use of IFs: Existing QoE models often rely on a narrow subset of IFs, typically objective metrics such as stall events, latency. This fragmented view overlooks important contextual and user-centric factors, which are critical for capturing the complexity of VVS.

V) Computational Complexity and Power Consumption: Processing large-scale 3D data and performing real-time adaptation impose substantial computational cost and power consumption on mobile and wearable devices, affecting battery life and thermals. Yet, many models still overlook efficiency and overhead, motivating energy-aware and lightweight QoE solutions that remain accurate while supporting real-time assessment.

VI) Multimodal QoE Models: Most current studies emphasizes visual quality, while real applications increasingly involve multimodal perception. Integrating audio, haptics, and spatial sound can better reflect realistic user experience and improve QoE modeling for emerging immersive scenarios.

V. Conclusion

VVS enables immersive 6DoF 3D, but sustaining high QoE is difficult because volumetric content is complex and network/device budgets are tight. We survey recent studies, covering subjective assessments and objective models grouped by IFs and methodology. Most models use few factors, focus on point clouds, and rarely account for power or other volumetric formats. Future models should be context-aware, work across formats, and lightweight for real-time mobile use.

VI. Acknowledgment

The research is supported in part by the Austrian Federal Ministry for Digital and Economic Affairs, National Foundation for Research, Technology and Development, and Christian Doppler Research Association with Project Christian Doppler Laboratory ATHENA (http://athena.itec.aau.at).Additionally, it was supported by project PID2022-140560OB-I00 (DRONAC), funded by MICIU/AEI/10.13039/501100011033, ERDF, and EU.

References

[1] J. v. d. Hooft *et al.*, "From Capturing to Rendering: Volumetric Media Delivery with Six Degrees of Freedom," *IEEE Comm. Mag.*, 2020.
[2] R. Pagés *et al.*, "Affordable Content Creation for Free-Viewpoint Video and VR/AR Applications," *Journal of Vis. Comm. and Image Rep.*, 2018.
[3] SNS Insider, "Volumetric Video Market Size, Share & Segmentation Report," 2022. [Online]. Available: https://is.gd/snsvvm
[4] Y. Liu *et al.*, "Vues: Practical Mobile Volumetric Video Streaming Through Multiview Transcoding," in *MobiCom*, 2022.
[5] K. Hu *et al.*, "Understanding User Behavior in Volumetric Video Watching: Dataset, Analysis and Prediction," in *ACM MM*, 2023.

[6] J. Li *et al.*, "Optimal Volumetric Video Streaming with Hybrid Saliency Based Tiling," *IEEE Transactions on Multimedia*, 2023.

[7] B. Li *et al.*, "DeformStream: Deformation-Based Adaptive Volumetric Video Streaming," 2024. [Online]. Available: https://is.gd/DSvvs

[8] K. Lee *et al.*, "GROOT: A Real-Time Streaming System of High-Fidelity Volumetric Videos," in *MobiCom*, 2020.

[9] C. Wang *et al.*, "VoLUT: Efficient Volumetric Streaming Enhanced by LUT-Based Super-Resolution," in *MLSys*, 2025.

[10] Y. Alkhalili *et al.*, "A Survey of Volumetric Content Streaming Approaches," in *IEEE BigMM*, 2020.

[11] Y. Shi *et al.*, "QV4: QoE-Based Viewpoint-Aware V-PCC-Encoded Volumetric Video Streaming," in *ACM MMSys*, 2024.

[12] A. Horé *et al.*, "Image Quality Metrics: PSNR vs. SSIM," in *International Conference on Pattern Recognition*, 2010.

[13] C. Jin *et al.*, "No-Reference Quality Assessment for 3D Synthesized Images Based on Visual-Entropy-Guided Multi-Layer Features Analysis," *Entropy*, 2021.

[14] E. Dumic *et al.*, "Point Cloud Coding Solutions, Subjective Assessment and Objective Measures: A Case Study," *Symmetry*, 2020.

[15] J. Li *et al.*, "A QoE Model in Point Cloud Video Streaming," 2022. [Online]. Available: https://arxiv.org/abs/2111.02985

[16] Y. Jin *et al.*, "From Capture to Display: A Survey on Volumetric Video," 2024. [Online]. Available: https://arxiv.org/abs/2309.05658

[17] R. B. Rusu *et al.*, "3D Is Here: Point Cloud Library (PCL)," in *IEEE ICRA*, 2011.

[18] G. Sandri *et al.*, "Compression of Plenoptic Point Clouds," *IEEE Transactions on Image Processing*, 2019.

[19] D. Bommes *et al.*, "Quad-Mesh Generation and Processing: A Survey," in *Computer Graphics Forum*, 2013.

[20] S. Schwarz *et al.*, "Emerging MPEG Standards for Point Cloud Compression," *IEEE JETCAS*, 2019.

[21] B. R. de Araújo *et al.*, "A Survey on Implicit Surface Polygonization," *ACM Computing Surveys*, 2015.

[22] B. Mildenhall *et al.*, "NeRF: Representing Scenes As Neural Radiance Fields for View Synthesis," *Communications of ACM*, 2021.

[23] B. Kerbl *et al.*, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Transactions on Graphics*, 2023.

[24] D. Le Gall, "MPEG: A Video Compression Standard for Multimedia Applications," *Commun. of the ACM*, vol. 34, no. 4, p. 46–58, 1991.

[25] J. M. Boyce *et al.*, "MPEG Immersive Video Coding Standard," *Proceedings of the IEEE*, 2021.

[26] H. M. Fornes *et al.*, "Acceptable Latency in Predictable First-Person VR Cloud Games," in *International Workshop on MMVE*, 2025.

[27] S. V. Damme *et al.*, "Impact of Latency on QoE, Performance, and Collaboration in Interactive Multi-User Virtual Reality," *Appl. Sci.*, 2024.

[28] K. Bouraqia *et al.*, "Quality of Experience for Streaming Services: Measurements, Challenges and Insights," *IEEE Access*, 2020.

[29] International Telecommunication Union (ITU-T), "Recommendation ITU-T P.1201: Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality," ITU-T, Recommendation P.1201, Oct. 2012. [Online]. Available: https://www.itu.int/rec/T-REC-P.1201/en

[30] N. Barman *et al.*, "QoE Modeling for HTTP Adaptive Video Streaming–A Survey and Open Challenges," *IEEE Access*, 2019.

[31] A. Raake *et al.*, "IP-Based Mobile and Fixed Network Audiovisual Media Services," *IEEE Signal Processing Magazine*, 2011.

[32] P. Le Callet *et al.*, "Qualinet White Paper on Definitions of Quality of Experience," Tech. Rep., Mar. 2013. [Online]. Available: https://theses.hal.science/EC-NANTES/hal-04638470v1

[33] P. Juluri *et al.*, "Measurement of Quality of Experience of Video-on-Demand Services: A Survey," *IEEE Comm. Surveys & Tutorials*, 2016.

[34] D. Nguyen *et al.*, "QoE Models for Adaptive Streaming: A Comprehensive Evaluation," *Future Internet*, 2022.

[35] L. Peroni *et al.*, "Empowerment of Atypical Viewers via Low-Effort Personalized Modeling of Video Streaming Quality," *Proceedings of ACM on Networking*, 2023.

[36] N. Wehner *et al.*, "A Tutorial on Data-Driven Quality of Experience Modeling with Explainable Artificial Intelligence," *IEEE Communications Surveys & Tutorials*, 2025.

[37] X. Yin *et al.*, "A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP," in *ACM SIGCOMM*, 2015.

[38] Z. Duanmu *et al.*, "A Bayesian Quality-of-Experience Model for Adaptive Streaming Videos," *ACM Tran. Mult. Comput. Comm. Appl.*, 2023.

[39] C. An *et al.*, "Enhancing QoE of Adaptive Video Streaming by Generating Fine-Grained Throughput," *IEEE TCSVT*, 2025.

[40] Z. Duanmu *et al.*, "Quality-of-Experience for Adaptive Streaming Videos: An Expectation Confirmation Theory Motivated Approach," *IEEE Transactions on Image Processing*, 2018.

[41] H. Zhang *et al.*, "DeepQoE: A Unified Framework for Learning to Predict Video QoE," in *IEEE ICME*, 2018.

[42] T. Huang *et al.*, "Optimizing Adaptive Video Streaming with Human Feedback," in *ACM MM*, 2023.

[43] A. Balachandran *et al.*, "Developing A Predictive Model of Quality of Experience for Internet Video," in *ACM SIGCOMM*, 2013.

[44] A. Zhang *et al.*, "Efficient Volumetric Video Streaming Through Super Resolution," in *International Workshop on HotMobile*, 2021.

[45] J. Li *et al.*, "Volumetric Video Streaming Subjective Experiment Database," 2022. [Online]. Available: https://is.gd/vvssed

[46] A. Zhang *et al.*, "YuZu: Neural-Enhanced Volumetric Video Streaming," in *USENIX NSDI*, 2022.

[47] X. Wang *et al.*, "Spatial Perceptual Quality Aware Adaptive Volumetric Video Streaming," in *IEEE GLOBECOM*, 2023.

[48] J. Weil *et al.*, "Modeling Quality of Experience for Compressed Point Cloud Sequences Based on A Subjective Study," in *QoMEX*, 2023.

[49] ——, "Point Cloud Sequence QoE Dataset," 2023. [Online]. Available: https://github.com/jw3il/point-cloud-qoe

[50] J. Li *et al.*, "Toward Optimal Real-Time Volumetric Video Streaming: A Rolling Optimization and Deep Reinforcement Learning Based Approach," *IEEE TCSVT*, 2023.

[51] ——, "SC-DDQN Dataset," 2023. [Online]. Available: https://github.com/FreezingWHY/SC-DDQN/tree/master

[52] Y. Gao *et al.*, "FRAS: Federated Reinforcement Learning Empowered Adaptive Point Cloud Video Streaming," 2023. [Online]. Available: https://arxiv.org/abs/2207.07394

[53] M. Nguyen *et al.*, "Characterization of the Quality of Experience and Immersion of Point Cloud Videos in Augmented Reality Through a Subjective Study," *IEEE Access*, 2023.

[54] ——, "QoE and Immersion of Dynamic Point Cloud," 2023. [Online]. Available: https://is.gd/qidpc

[55] D. V. Nguyen *et al.*, "Modeling User Quality of Experience in Adaptive Point Cloud Video Streaming," in *Inter. Symp. on Multimedia*, 2024.

[56] ——, "Dynamic Point Cloud Subjective Test," 2024. [Online]. Available: https://is.gd/Dpcst

[57] Y. Shi *et al.*, "Quality Assessment and Modeling for MPEG V-PCC Volumetric Video," in *International Workshop on MMVE*, 2024.

[58] K. Hu *et al.*, "LiveVV: Human-Centered Live Volumetric Video Streaming System," *IEEE Internet of Things Journal*, 2025.

[59] A. Bentaleb *et al.*, "Solutions, Challenges and Opportunities in Volumetric Video Streaming: An Architectural Perspective," *ACM Transactions on Multimedia Computing, Communications and Applications*, 2024.

[60] J. v. d. Hooft *et al.*, "Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming," in *QoMEX*, 2020.

[61] S. Subramanyam *et al.*, "Subjective QoE Evaluation of User-Centered Adaptive Streaming of Dynamic Point Clouds," in *QoMEX*, 2022.

[62] C. He *et al.*, "COVER: A Comprehensive Video Quality Evaluator," in *IEEE/CVF Conference on CVPRW*, 2024, pp. 5799–5809.

[63] E. Zerman *et al.*, "Textured Mesh vs Coloured Point Cloud: A Subjective Study for Volumetric Video Compression," in *QoMEX*, 2020.

[64] E. S. Jang *et al.*, "Video-Based Point-Cloud-Compression Standard in MPEG: From Evidence Collection to Committee Draft [Standards in A Nutshell]," *IEEE Signal Proc. Mag.*, vol. 36, no. 3, pp. 118–123, 2019.

[65] ISO/IEC, "Information Technology — Coded Representation of Immersive Media - Part 9: Geometry-Based Point Cloud Compression," ISO/IEC, International Standard ISO/IEC 23090-9:2023, Mar. 2023. [Online]. Available: https://www.iso.org/standard/78990.html

[66] A. Takahashi *et al.*, "Standardization Activities in the ITU for A QoE Assessment of IPTV," *IEEE Communications Magazine*, 2008.

[67] J. Nightingale *et al.*, "5G-QoE: QoE Modelling for Ultra-HD Video Streaming in 5G Networks," *IEEE Transactions on Broadcasting*, 2018.

[68] L. Wang *et al.*, "QoE-Driven Adaptive Streaming for Point Clouds," *IEEE Transactions on Multimedia*, 2023.

[69] S. R. Cox *et al.*, "VOLVQAD: An MPEG V-PCC Volumetric Video Quality Assessment Dataset," in *ACM MMSys*, 2023.

[70] A. Clemm *et al.*, "Toward Truly Immersive Holographic-Type Communication: Challenges and Solutions," *IEEE Comm. Magazine*, 2020.

[71] Z. Liu *et al.*, "Point Cloud Video Streaming: Challenges and Solutions," *IEEE Network*, 2021.

[72] N. Eswara *et al.*, "Streaming Video QoE Modeling and Prediction: A Long Short-Term Memory Approach," *IEEE TCSVT*, 2020.