Signal Processing: Image Communication 143 (2026) 117514

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Quality assessment of 3D reconstructed meshes: Bridging objective metrics, subjective perception, and behavioral cues ☆

Anna Ferrarotti [a], Isabel Rodríguez [b], Javier Usón [b], Sara Baldoni [c,*], David Barbero [b], Daniel Berjón [b], Francisco Morán [b], Narciso García [b], Federica Battisti [c], Jesús Gutiérrez [b], Marco Carli [a], Julián Cabrera [b]

[a] Roma Tre University, Rome, Italy
[b] Grupo de Tratamiento de Imágenes, Information Processing and Telecommunications Center (IPTC), ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain
[c] University of Padova, Padua, Italy

ARTICLE INFO

Dataset link: https://github.com/gti-upm/CH3D-Reco

Keywords:
3D reconstruction
Quality assessment
Interactive media
Cultural heritage

ABSTRACT

Assessing the quality of 3D reconstructed models remains a key challenge in multimedia applications, especially in the context of cultural heritage, where visual fidelity and perceptual realism are equally crucial. This study investigates how reconstruction parameters, as well as existing objective quality metrics, align with human perception. In addition, we analyze how perceived quality and user interaction are related. A dataset of 3D models was generated by varying the number of input images, mesh complexity, and texture resolution. Results from a subjective study show that texture resolution significantly affects perceived quality, whereas variations in number of images and mesh complexity have a limited impact. Furthermore, interaction behavior was found to vary with perceived quality, with participants spending more time and exploring larger viewing angles for models receiving higher scores. These findings highlight the need for perceptually grounded, interaction-aware evaluation methodologies and provide guidelines for future perceptual optimization of 3D reconstruction pipelines.

1. Introduction

The widespread diffusion of 3D content has transformed multimedia research and applications, ranging from gaming [1] to e-commerce [2], telepresence systems [3], and cultural heritage [4]. Based on this, the need for realistic, interactive, and efficient 3D representations is growing. Central to this progress is the ability to acquire and reconstruct high-quality 3D models, which serve as the foundation for visualization, interaction, and analysis across domains. However, the lack of an evaluation metric that relates rendering parameters and the quality of 3D models limits the understanding of users' preferences, whose satisfaction plays a decisive role in enabling meaningful multimedia experiences [5].

Due to the growing interest in the field, 3D reconstruction techniques have advanced considerably, with methods such as Structure from Motion (SfM) and Multi-View Stereo (MVS) [6] providing scalable pipelines for generating detailed polygonal meshes from images. More recently, other methodologies, such as Neural Radiance Fields (NeRFs) [7] and Gaussian Splatting (GS) [8], have demonstrated impressive photorealism, though at the cost of limited compatibility with conventional 3D pipelines. Despite such progress, the evaluation of reconstruction quality remains an open challenge. Objective metrics have been proposed to automatically assess quality, yet they often fail to align with human visual perception. On the other hand, subjective evaluations, though reliable, are time- and resource-intensive. Identifying a trade-off between scalability and perceptual validity is central to multimedia research.

The challenge is further compounded by the fact that reconstructed 3D models are typically consumed in interactive and heterogeneous environments, where factors such as display devices, rendering conditions, and user engagement may influence perceived quality. In multimedia applications, this creates a dual requirement: reconstructions must remain faithful to the original content while being optimized for transmission, storage, and real-time rendering. Understanding how reconstruction parameters – such as mesh resolution or texture fidelity

☆ This article is part of a Special issue entitled: 'IMAGE_Immersive and Networked Media' published in Signal Processing: Image Communication.

* Corresponding author.
E-mail address: sara.baldoni@unipd.it (S. Baldoni).

https://doi.org/10.1016/j.image.2026.117514
Received 15 October 2025; Received in revised form 4 February 2026; Accepted 8 February 2026
Available online 10 February 2026

– affect perceptual quality is therefore critical for developing adaptive pipelines and ensuring optimal user experience.

Within this broad multimedia landscape, cultural heritage represents a particularly compelling use case. Accurate 3D reconstructions of historical objects, monuments, and historical sites enable preservation [9], documentation [10], and dissemination [11] at unprecedented scales. These models can be employed in museum exhibitions, virtual tours, educational tools, and restoration projects, where their quality directly impacts both scholarly utility and public engagement [12]. Cultural heritage applications also highlight the interplay between technical fidelity and subjective perception: while fine details are essential for conservators and researchers, visual plausibility and interactivity are equally critical for engaging wider audiences. This duality makes cultural heritage an ideal testbed for exploring how objective metrics align with human perception and how interaction shapes quality judgments.

In this study, we address these challenges by focusing on three research questions:

RQ1: How does perceived quality vary with reconstruction parameters such as the number of input images used, mesh complexity, and texture resolution?

RQ2: Are existing objective metrics correlated with human visual perception when applied to 3D reconstructions?

RQ3: Is user interaction driven by model quality?

By systematically investigating these questions, we aim to bridge the gap between quantitative measures and user-centered evaluations, providing new insights into the assessment of 3D model quality. While the experimental framework is developed within the cultural heritage context, the findings are broadly applicable to multimedia applications where the perceptual quality of 3D content is critical. Besides addressing the selected research questions, information about the complete dataset, including the 3D models, subjective scores, and objective evaluations, is available at https://github.com/gti-upm/CH3D-Reco.

The remainder of this work is organized as follows. Section 2 discusses the current state of the art both for 3D model reconstruction and quality evaluation. Section 3 outlines the methodology employed to answer our research questions. In Sections 4 and 5, the results are reported and discussed, respectively. Lastly, Section 6 draws the conclusions of this study.

2. Related works

3D reconstruction is a rapidly evolving research field, though it builds upon well-established and widely adopted methods. Reconstruction approaches can be broadly classified into active and passive methods [13]. Active methods employ dedicated hardware to directly scan object surfaces. A common example is laser scanning [14], where object geometry is derived from measurements of reflected laser beams, often combined with LiDAR data. While effective, these techniques require specialized equipment and are typically costly. In contrast, passive methods rely on image acquisition devices to capture the radiance reflected by the surfaces. The most widely used technique in this category is SfM, which reconstructs a 3D structure from images acquired from multiple viewpoints [15]. SfM estimates both camera motion and object geometry by tracking key features across images, and it is particularly effective when camera trajectories are well defined and the scene is not significantly affected by occlusions or complex lighting [13]. To improve density and accuracy, MVS is commonly applied after SfM. Other established approaches include Shape from Shading (SfS) and Shape and Albedo from Shading (SAfS), which infer shape – or both shape and reflectance – by analyzing shading cues [16,17].

Recent research has increasingly focused on two directions: deep learning-based methods and GS [18]. Among the former, NeRF has gained significant attention for its ability to generate high-resolution reconstructions, though at substantial computational cost [19]. Other approaches leverage Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs): CNNs are versatile and can reconstruct models even from a single image, but may produce less detailed models [20], while GANs excel at producing highly photorealistic results [21]. However, both require heavy computational resources. To mitigate this, Variational Autoencoders (VAE)-based methods have been proposed [22], offering reduced storage requirements and faster computation, at the expense of a loss in reconstruction detail. Hybrid approaches combining VAEs with GANs attempt to balance these trade-offs [23].

More recently, GS has emerged as a promising alternative, enabling highly photorealistic renderings without deep learning architectures [24]. In this approach, the object is represented as a set of 3D Gaussians modeling its geometry and appearance. Despite its advantages, this method is still affected by artifacts such as floating elements, which can compromise rendering fidelity [13].

While neural methods such as NeRF and GS achieve superior photorealism, their reliance on learned representations hinders integration with conventional 3D software pipelines. Mesh-based methods, by contrast, provide a practical and widely compatible solution. For this reason, polygonal meshes, obtained by the combination of SfM and MVS, remain the prevailing standard for representing 3D geometry in modeling, editing, and visualization. Moreover, photogrammetry – one of the most widely adopted techniques in the cultural heritage domain – relies on SfM as a fundamental processing step for generating accurate 3D models. However, existing research on 3D reconstruction has largely concentrated on algorithmic accuracy and photorealism, while the relationship between variations in reconstruction pipelines and parameters and the perceived quality of the resulting 3D meshes has received limited attention.

Furthermore, despite continuous advances in reconstruction techniques, there is still no broad consensus on how to evaluate the quality of the resulting 3D models. Quality assessment is critical for ensuring that reconstructions meet the requirements of their intended applications — for example, guaranteeing fidelity in cultural heritage preservation or enabling efficient optimization for storage and transmission.

Quality can be assessed either through subjective or objective methods. Subjective evaluation relies on user studies, where participants rate the quality of reconstructions under different conditions. ITU-T P.910 [25] and ITU-T P.919 [26] describe standardized methodologies for video quality assessment in conventional and 360° settings, respectively. Among these, Absolute Category Rating (ACR) and Degradation Category Rating (DCR) are the most commonly employed in the literature. In ACR, participants rate a single presented stimulus, whereas in DCR two stimuli are shown: the reference and its distorted version. These methodologies are frequently adopted for 3D model evaluation as well. ACR is often used for exploratory analyses in immersive environments [27,28], while DCR typically yields more statistically robust results and is regarded as more reliable [29].

In [30], a subjective study on 3D meshes employed the DCR method: participants viewed side-by-side presentations of a reference mesh and its distorted counterpart, rendered as 8-second rotation videos, and rated the impairment on a five-level scale ranging from imperceptible to very annoying. The DCR procedure was also adopted in [31]. Five textured 3D meshes were simplified into four Levels of Detail (LoDs) using edge contraction with quadric error metrics. In a Virtual Reality (VR) environment, participants compared each LoD mesh to its reference while both rotated three times. After each trial, participants rated the impairment on a five-level scale. Twenty participants completed the study with sessions lasting about 30 min. In [32], the evaluation focused on NeRF-based view synthesis. Stimuli consisted of 88 video pairs (reference vs. synthesized, covering front-facing and 360° scenes) displayed side by side in randomized order. Subjects rated quality on

(a) Horn of Salt Diggers Brotherhood of Wieliczka. (b) Palace of Fine Arts. (c) Mexico City Metropolitan Cathedral.

(d) Roman Temple of Evora. (e) Kriegerdenkmal.

Fig. 1. Selected models for objective metrics performance evaluation.

a continuous 0–100 scale with five labeled categories: Bad, Poor, Fair, Good, and Excellent.

Objective evaluation, by contrast, relies on computational metrics designed to correlate with human perception. These can be categorized as Full Reference (FR) [33–37], Reduced Reference (RR) [38,39], or No Reference (NR) [40–44], depending on whether a reference model is used during the metric computation. Since original models are usually missing when reconstructing 3D content from real objects, NR approaches are the most suitable. These metrics can also be classified as either model-based or projection-based, depending on whether features are extracted directly from the 3D mesh or from several 2D renderings extracted from different viewpoints.

Several NR methods have been proposed in the literature. For instance, [40] introduced a saliency-driven approach: a saliency map is computed on the 3D mesh, and 2D renderings are generated from multiple viewpoints. The most salient views are selected, divided into patches, and processed by three CNN backbones (AlexNet, VGG, ResNet). The resulting feature vectors are fused using Compact Multilinear Pooling (CMP), yielding a reduced-dimensional representation from which the final quality score is predicted.

Zhang et al. [41] proposed a quality metric applicable to both 3D meshes and point clouds, inspired by the Natural Scene Statistics (NSS) paradigm from image processing. Geometric features (curvature, dihedral angle, face area, and angle) and color features (in the LAB color space) are extracted, and a set of statistical descriptors, such as mean, standard deviation, entropy, and parameters of generalized Gaussian, asymmetric Gaussian, and gamma distributions, are computed. These features are then fed into a Support Vector Regressor (SVR) to predict quality.

Projection-based methods have also been explored to leverage mature 2D image quality assessment techniques. [42] introduced EEP-3DQA, a NR metric designed to improve the efficiency of previously proposed methods. Following the MPEG VPCC setup [45], six cube-face projections of the 3D model are generated, and a random subset of these views is sampled to reduce redundancy. A grid mini-patch sampling strategy further organizes each projection into patches, which are processed by a lightweight Swin-Transformer tiny backbone, followed by a two-stage fully connected regressor. In [43], a similar pipeline was proposed, in which all the projections are retained, and random sampling is performed at the patch level rather than at the projection level, allowing efficiency optimization while preserving completeness of the visual information.

Finally, the authors in [44] proposed to model 3D meshes as graphs. In this representation, mesh vertices and edges are mapped into a weighted graph structure, with edge weights derived from LAB color features. From this graph, several topological descriptors (e.g., vertex degree, vertex strength, centrality measures) are extracted, and their statistical moments (mean, variance, skewness, kurtosis, entropy) are computed. These descriptors form the feature set for a Random Forest Regressor (RFR) used to estimate perceived quality.

Despite the large number of metrics proposed in the literature, their practical use remains highly challenging. Indeed, while modern metrics often demonstrate excellent performance on the specific datasets used for training and validation, their ability to generalize to new models – such as those obtained from 3D reconstructions – is limited. At the same time, their capability to provide quality assessments consistent with human perception is still an open question. In addition, many of these metrics rely on machine learning or deep learning algorithms, and their source code is often not made publicly available. Therefore, reproducing the results reported in the original studies can be complex. Considering these limitations, subjective experiments, although costly and time-consuming, remain essential for reliably assessing the perceived quality of 3D models.

From a perceptual perspective, it is also important to note that the outcome of 3D reconstruction pipelines depends on several algorithm-specific parameters, such as the number of input images in SfM or the complexity of the final mesh. However, to the best of our knowledge, a systematic study investigating how user perception varies with reconstruction parameters (e.g., input image count, mesh complexity, or texture resolution) is lacking. Similarly, no prior work assessed how well commonly used objective mesh quality metrics reflect such perceptual changes.

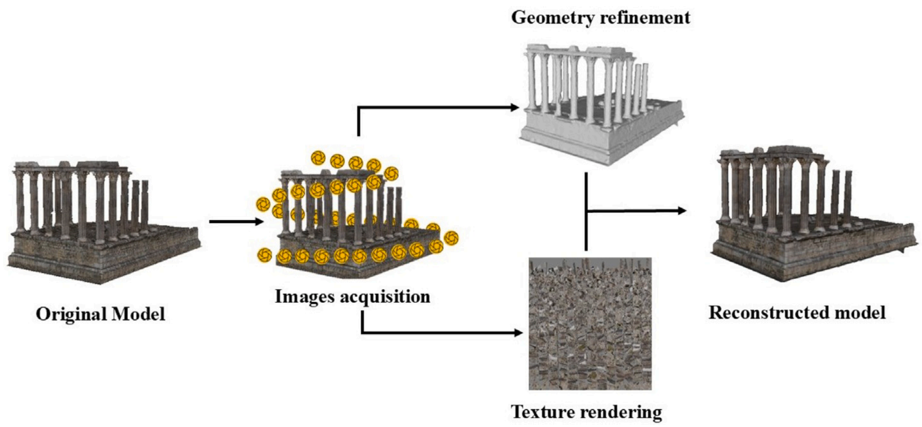


Fig. 2. Method employed for 3D model reconstruction using SfM. Captures of the model are taken from 1000 different positions, then a subset of such images are used to reconstruct the geometry as a triangular 3D mesh and to create a texture atlas for the model.

Moreover, although most existing studies assess quality under passive viewing conditions, many contemporary multimedia applications – including cultural heritage visualization – are inherently interactive. The effect of user interaction on perceived 3D quality, and its relationship with both reconstruction parameters and objective metrics, remains an open research question. This work addresses these gaps by systematically analyzing subjective perception of 3D mesh quality under controlled variations of reconstruction parameters, comparing human judgments with widely used objective quality metrics, and investigating the influence of interactivity on perceived quality.

3. Proposed method

This section presents the methodology adopted to address the three research questions introduced in Section 1. The first phase of the study focused on generating a dataset of 3D models, which was then used both to evaluate objective quality metrics and to conduct a subjective experiment aimed at assessing their correlation with human perception (**RQ1** and **RQ2**). The 3D models were reconstructed following the SfM and MVS pipeline, which is currently the most widely employed approach, including in commercial software, as mentioned in Section 2. To specifically address **RQ1**, the models were reconstructed under different parameter configurations, such as the number of input images, the number of mesh triangles, and the texture resolution, as detailed in Section 3.1. For **RQ3**, the application used during the subjective experiment was designed to enable users to interact with the models before providing their quality ratings. Further details are provided in Section 3.2. Finally, Section 3.3 provides an overview of the participants involved in the study.

3.1. 3D models selection and generation

The original 3D models were selected from the BASICS dataset [46], which contains a broad collection of point clouds organized into three semantic categories: (i) buildings and landscapes, (ii) inanimate objects, and (iii) humans and animals. The point clouds in the dataset were derived from 3D meshes available on Sketchfab,[1] which serve as the source models for this study. Since the focus of this work is on cultural heritage applications, five source models, shown in Fig. 1, were chosen from the first two BASICS categories. These models were then used to simulate the 3D reconstruction process, as illustrated in Fig. 2 and described in detail in the following sections.

This pipeline was briefly introduced in our previous work [47], and it is detailed here for the sake of completeness and clarity. For each source model, we generated 1000 synthetic captures using Blender[2]. The virtual camera was configured with a fixed field of view of 42°, and its positions were uniformly distributed over the surface of a sphere centered on the model. This spherical distribution ensured comprehensive coverage of the model's exterior geometry, thereby reducing the risk of reconstruction artifacts such as missing surfaces or holes. To achieve consistent sampling, the captures were organized along a spiral trajectory extending from the base to the uppermost point of the model. Each capture was rendered at a resolution of 3840 × 2160 pixels, and the full set of intrinsic and extrinsic camera calibration parameters was recorded to facilitate accurate reconstruction.

Reconstruction was performed using the open-source SfM and MVS toolchain comprising OpenMVG [48] and OpenMVS [49]. OpenMVG was employed to estimate camera parameters and to generate an initial sparse point cloud. Subsequently, OpenMVS applied MVS algorithms to compute depth maps from the input images, which were then fused into a dense point cloud. This dense point cloud was converted into a triangular surface representation via Poisson Surface Reconstruction [50]. To control model complexity, the resulting meshes were further simplified using Quadric Edge Collapse Decimation, enabling systematic variation in mesh resolution. Finally, surface appearance was reconstructed by applying texture mapping using the approach described in [51].

The advantages of using this simulation are twofold. First, relying on publicly available models and state-of-the-art reconstruction methods ensures that the results of this study can be easily replicated. Second, employing models with an available reference makes it possible to apply FR metrics.

To systematically investigate the influence of reconstruction parameters, we produced 64 reconstructed variants of each source model by varying three factors:

- number of input images provided for reconstruction – 50, 125, 250, or 500;
- number of triangles in the reconstructed mesh – 5000, 25 000, 50 000, or 100 000;
- texture resolution – 1, 4, 16, or 32 Mega Pixels (MPs).

A preliminary study involving a focus group of five experts was conducted to assess reconstructed models generated using a pool of reconstruction parameter settings. Based on their evaluations, the parameters for the main subjective experiment were selected to ensure perceptually distinguishable quality levels while maintaining a balance between reconstruction fidelity and computational cost.

[1] https://sketchfab.com.

[2] https://www.blender.org.

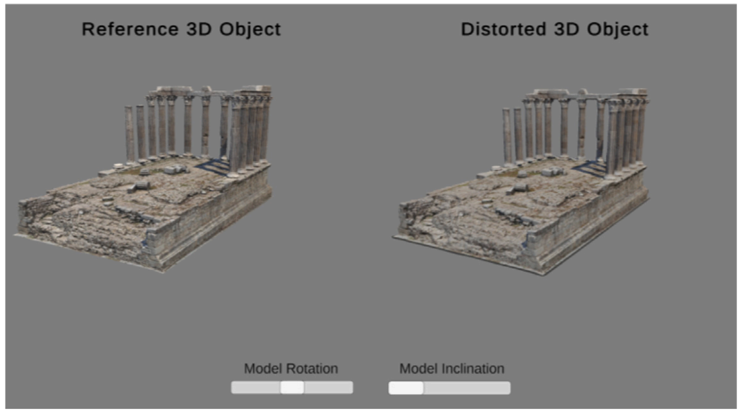


Fig. 3. Interactive interface used for the subjective evaluations.

3.2. Experimental protocol

To collect user quality ratings, the DCR method was adopted. By directly comparing the original models available on Sketchfab and the reconstructed models, participants were able to assess the visual quality of the 3D reconstructions with higher accuracy and consistency. Furthermore, as discussed in Section 2, DCR is generally considered more reliable than ACR.

An interactive evaluation interface was developed in Unity. During the experiments, participants viewed on a desktop PC a pair of 3D models, the reference one and its reconstructed version, as shown in Fig. 3. Model pairs were selected through a randomization algorithm that avoided repetitions and prevented consecutive evaluations of the same model. Since one of the aims of this study is to analyze the effects of interaction on the perceived quality, we introduced a modified version of the standard DCR protocol, in which users could freely explore the models, instead of being provided with pre-rendered videos. The application synchronized the rotation and tilting of both models. The viewing perspective in terms of inclination and rotation was controlled through sliders, ensuring that both models were always viewed from identical perspectives.

After completing the visual inspection, participants proceeded to a voting window where they rated the quality of the reconstructed model relative to the original on a 5-point Likert scale. All results were automatically stored in CSV format. The application also logged all interactions (rotations and tilts) along with their timestamps.

The experimental protocol consisted of two phases: a training session and a test session. The training phase allowed participants to become familiar with both the exploration and rating procedures. In this phase, three distorted versions of a 3D model (not included in the test session) were used. Before starting the test phase, participants provided basic demographic and contextual information, including an identification code, age, gender, the assigned playlist to be evaluated, i.e., the specific sequence of reconstructed models, and their prior experience with 3D models. During the test session, each participant was asked to rate two different playlists, each containing a distinct subset of 3D models. Participants were allowed unlimited time for evaluation and could take a short break between playlists to reduce visual fatigue and ensure consistent performance throughout the session.

3.3. Participants

In accordance with the GDPR [52], all participants were informed about the data collection procedure and signed a written informed consent form prior to their participation in the study. 41 participants (24 men and 17 women) took part in the experiment. Their ages ranged from 20 to 70 years (27.61 ± 11.35). Specifically, 82.93% of the participants were between 20 and 30 years old, 9.76% were between 31 and 50, and 7.32% were older than 51 years. Regarding prior experience with 3D models, 63.41% of the participants reported having used 3D models fewer than 5 times, 9.76% between 5 and 20 times, and 26.83% more than 20 times.

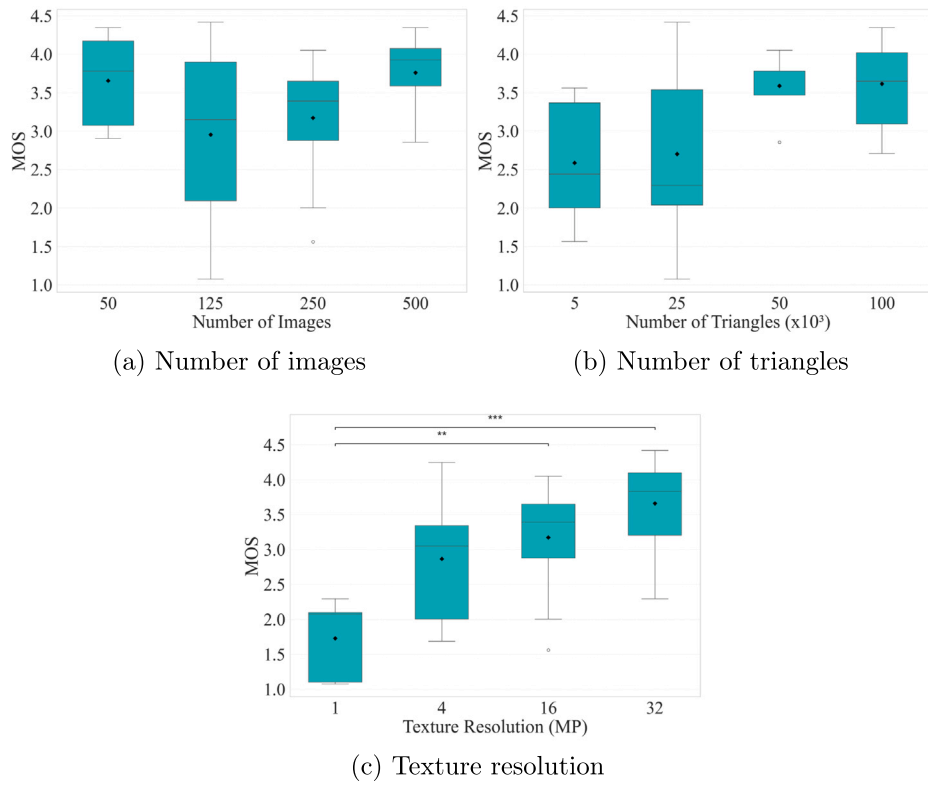
4. Results

In this section, we detail the results of the performed study, thus addressing the identified research questions.

4.1. Variation of subjective perception with reconstruction parameters (RQ1)

To evaluate the influence of the reconstruction parameters on perceived quality (**RQ1**), the MOS was initially computed for each reconstructed model as the average of the scores assigned by the participants. To determine whether statistically significant differences were present among the different parameter values used for model reconstruction (number of images, number of triangles, and texture resolution), the data were first tested for compliance with the assumptions of normality and homoscedasticity required for conducting a one-way Analysis of Variance (ANOVA) test [53]. These assumptions were verified using the Kolmogorov–Smirnov [54] and Levene [55] tests, respectively. The ANOVA tests were carried out with a significance level of 5%; accordingly, the null hypothesis was rejected for p-values lower than 0.05. When the ANOVA indicated significant differences among reconstruction levels, post-hoc pairwise comparisons were performed using Tukey's Honestly Significant Difference (HSD) test [56]. When the assumptions of the ANOVA were not satisfied, the Kruskal–Wallis H test [57] was used to verify significant differences among the groups, followed by a Mann–Whitney U test with Bonferroni correction for pairwise comparisons [58], using also in this case a 5% significance level. Fig. 4 illustrates the distribution of the MOS values as a function of the reconstruction parameters, while Table 1 reports the detailed p-values obtained from the statistical analyses.

With respect to the number of images, the ANOVA test did not reveal statistically significant differences. In addition, as shown in Fig. 4(a), no particular trend was observed with respect to the variation in the number of images used for reconstruction. In contrast, both the number of triangles and the texture resolution exhibited an increasing trend (Figs. 4(b) and 4(c)). In both cases, the null hypothesis was rejected by the ANOVA and the Kruskal–Wallis tests, indicating significant differences among the groups. However, concerning the number of triangles, although Fig. 4(b) shows a slight increasing trend when the number of triangles grows, the pairwise comparisons failed to identify significant differences between the groups. On the other hand, Fig. 4(c) clearly shows an increasing trend of the MOS values with higher megapixel resolutions. Tukey's HSD test identified significant pairwise differences between 1 and 16 MP, and between 1 and 32 MP.



(a) Number of images (b) Number of triangles

(c) Texture resolution

Fig. 4. MOS distribution with respect to 3D models reconstruction parameters. The asterisks mark significance leveles: '*' for $p < 0.05$, '**' for $p < 0.01$, '***' for $p < 0.001$.

Table 1
Results of the statistical analysis performed on the MOS of the 3D models when varying the reconstruction parameters. A '–' indicates that the test was not performed.

Factor	Test performed	ANOVA p-value	Parameter comparison	Tukey HSD p-value
Number of images	ANOVA	0.193	50 vs. 125	–
			50 vs. 250	–
			50 vs. 500	–
			125 vs. 250	–
			125 vs. 500	–
			250 vs. 500	–
Number of triangles	Kruskal–Wallis	0.017*	5 vs. 25	1.0
			5 vs. 50	0.356
			5 vs. 100	0.095
			25 vs. 100	0.695
			25 vs. 50	0.086
			50 vs. 100	1.0
Texture resolution	ANOVA	<0.001*	1 vs. 4	0.063
			1 vs. 16	0.001*
			1 vs. 32	<0.001*
			4 vs. 16	0.831
			4 vs. 32	0.121
			16 vs. 32	0.188

* Highlights a statistically significant result.

4.2. Correlation between subjective perception and objective metrics (RQ2)

To address **RQ2**, a set of objective metrics was first selected to evaluate their correlation with the subjective scores collected during the user study. Specifically, four FR image quality metrics were considered (Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), Fréchet Inception Distance (FID)), together with two NR image quality metrics (Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and Naturalness Image Quality Evaluator (NIQE)), and two metrics directly related to the assessment of 3D model quality (Hausdorff distance and L2 distance), both FR. With the exception of FID, all the considered metrics are widely used in the field of 3D model quality assessment [59–61]. FID, originally introduced in the context of generative modeling [62], is here explored as a novel solution for quality assessment. Although generative and reconstruction-based scenarios differ, FID captures distribution-level perceptual characteristics that may provide complementary insights when applied to reconstructed 3D models.

Regarding the methodology adopted for quality evaluation through Image Quality Assessment (IQA) metrics, a total of 300 2D captures per model were used; these images were not employed during the reconstruction process.

Additionally, since the reconstructed models often exhibit variations in orientation, scale, and position, the camera calibration parameters obtained from the SfM process were compared with the original camera parameters. This comparison enabled the reconstructed models to be rotated, scaled, and translated to align them with the corresponding reference models, thereby reducing the overall complexity of computing both 2D and 3D quality metrics. For each 2D FR metric, comparisons were performed using captures acquired from identical viewpoints for both the reference and reconstructed models. In both cases, the backgrounds of the captures contained pixel values irrelevant to the metric computations; therefore, a binary mask including only the relevant pixels was generated for each capture. As a result of the reconstruction process, the masks corresponding to the same viewpoint in the reference and reconstructed models did not perfectly coincide. To address this, their union was computed, following the approach proposed in [47].

To evaluate the correlation between the objective and subjective metrics, both the Pearson correlation coefficient and the Spearman rank correlation coefficient were computed, revealing potential linear and monotonic relationships, respectively [53]. For each metric, the corresponding correlation coefficient, associated p-value, and coefficient of determination (R^2) were computed with respect to the subjective scores. The correlation coefficient (r and ρ, for Pearson's and Spearman's tests, respectively) measures the strength and direction of the relationship between two variables, while the coefficient of determination (R^2) represents the proportion of variance in the subjective scores that can be explained by the objective metric. As in the previous analyses, all statistical tests were performed using a significance level of 5%.

The results are reported in Table 2. Except for BRISQUE, all the analyses yielded p-values lower than 0.05, leading to the rejection of the null hypothesis of no correlation. However, NIQE, Hausdorff, and L2 distances exhibit weak correlations and explain only a limited portion of variance. On the other hand, PSNR and SSIM account for approximately 50% of variance and display considerably stronger correlation coefficients, ranging around 0.720. Among the metrics usually employed for 3D quality assessment, the best-performing one is LPIPS, which shows a strong negative correlation ($r = -0.789$) and explains around 63% of variance. Interestingly, FID also demonstrates a strong and statistically significant negative correlation with subjective quality scores ($r = -0.756$ and $\rho = -0.868$), explaining more than 57% and 75% of the variance for Pearson's and Spearman's coefficients, respectively. Although FID was originally introduced for evaluating generative models and is not specifically designed for 3D reconstruction tasks, this result suggests that distribution-based perceptual representations can capture aspects of perceived quality that are consistent with human judgments, even in reconstruction-based scenarios.

Overall, these results indicate that FR metrics are able to capture perceptually relevant differences in 3D reconstructions. However, FR applicability in real-world reconstruction scenarios is limited by the lack of reference models. Concurrently, image-based NR measures have demonstrated poor performance with respect to their consistency with subjective ratings, highlighting the need for alternative NR metrics. In this direction, the reliability of the FR measures demonstrated by the present study makes them a valuable benchmark for the future development of perceptually reliable NR metrics.

4.3. Link between interaction and quality (RQ3)

To address **RQ3**, we analyzed whether the quality of the reconstructed 3D models influenced users' interaction behavior. Specifically, we investigated how users explored the 3D content both temporally (time spent during exploration) and spatially (viewing angle).

As a first analysis, we examined how users' perceived quality was related to the time spent evaluating each 3D model. More specifically, we assessed whether time varied across the subjective scores assigned by the participants. Following the same statistical procedure described in Section 4.1, since the data did not satisfy the assumption of homoscedasticity, the non-parametric Kruskal–Wallis test was applied. The test led to the rejection of the null hypothesis (p-value < 0.001), indicating a significant difference in exploration time across the perceived quality levels. For pairwise comparisons, the Mann–Whitney U test with Bonferroni correction was applied, and the results are reported in Table 3, while Fig. 5 shows the distribution of exploration times across the subjective scores.

Table 2
Correlation analysis between subjective scores and objective metrics.

Metric	Pearson			Spearman		
	r	p-value	R^2	ρ	p-value	R^2
PSNR	0.721	<0.001*	0.520	0.713	<0.001*	0.508
SSIM	0.718	<0.001*	0.516	0.694	<0.001*	0.482
LPIPS	−0.795	<0.001*	0.633	−0.749	<0.001*	0.561
BRISQUE	−0.009	0.952	<0.001	0.124	0.415	0.015
NIQE	0.363	0.014*	0.132	0.413	0.005*	0.170
Hausdorff distance	−0.314	0.036*	0.099	−0.408	0.005*	0.167
L2 distance	−0.520	<0.001*	0.270	−0.556	<0.001*	0.309
FID	−0.756	<0.001*	0.571	−0.868	<0.001*	0.754

* Indicates a statistically significant result.

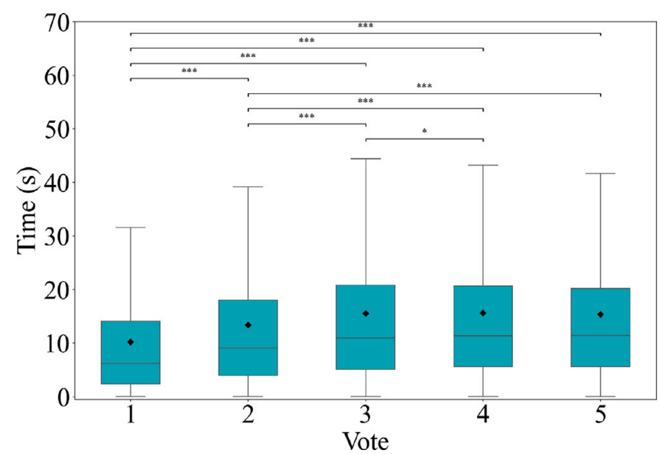


Fig. 5. Exploration time with respect to provided subjective scores. Outliers were removed for visualization purposes. The asterisks mark significance levels: '*' for $p < 0.05$, '**' for $p < 0.01$, '***' for $p < 0.001$.

Interestingly, the time spent by users differed significantly between low- and high-scoring models. Specifically, participants spent more time examining higher quality models. However, between the highest quality scores, 3 vs. 5 and 4 vs. 5, the differences in exploration time were no longer statistically significant.

To further investigate temporal variations in user interaction, we examined whether the exploration time was influenced by the reconstruction parameters used for generating the 3D models. Accordingly, statistical analyses were performed by considering the exploration time with respect to the different reconstruction parameters. In this case, since the data satisfied the assumptions of normality and homoscedasticity, a one-way ANOVA test was applied. However, no significant differences were found for the exploration time with respect to reconstruction parameters.

To assess whether quality evaluation was also influenced by the spatial exploration of the model, we analyzed how users rotated the 3D content during interaction. In particular, the rotation (yaw) angle was considered, and the same statistical tests used for exploration time were applied to the total angular displacement, the mean angle, and the angle variance. The total angular displacement is calculated as the sum of the absolute angular differences between each pair of consecutive rotation samples, expressed in degrees, i.e., the cumulative amount of

Table 3
Pairwise comparisons among subjective quality scores for the exploration time.

Votes	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5	2 vs. 3	2 vs. 4	2 vs. 5	3 vs. 4	3 vs. 5	4 vs. 5
p-value	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	0.027*	1.000	1.000

* Indicates a statistically significant result.

Table 4
Pairwise comparisons among subjective quality scores (in the range 1 to 5) for spatial exploration.

Angular feature	Vote									
	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5	2 vs. 3	2 vs. 4	2 vs. 5	3 vs. 4	3 vs. 5	4 vs. 5
Total displacement	<0.001*	<0.001*	<0.001*	<0.001*	0.543	<0.001*	0.024*	0.132	0.980	1.000
Mean	1.000	0.043*	0.248	1.0	0.092	0.720	1.0	1.0	1.0	1.0
Variance	0.134	0.004*	<0.001*	0.004*	0.735	0.066	0.484	0.483	0.951	0.994

* Indicates a statistically significant result.

rotation over time, regardless of direction, which provides an indication of how much the viewpoint moved during the trial.

Based on the results of the normality and homoscedasticity tests, the Kruskal–Wallis test was performed for the total angular displacement and the mean angle, whereas for the angle variance it was possible to apply the ANOVA test. All tests led to the rejection of the null hypothesis, with the following p-values: $p < 0.001$ for the total angular displacement, $p = 0.019$ for the mean angle, and $p < 0.001$ for the angle variance. Table 4 reports the p-values obtained from the pairwise comparisons, while Fig. 6 shows the distribution of the values across different votes. The results indicate a significant difference in total angular displacement, particularly between lower and higher quality scores, whereas the differences tend to diminish for the comparisons between higher subjective scores (above 3). The total displacement generally increases with perceived quality. Also, angle variance shows statistically significant differences between the score 1 and all the scores above 3, but no other comparisons yielded statistically significant results. Conversely, the mean angle does not exhibit substantial significant differences, except for the comparison between scores 1 and 3.

Given the results obtained for the total angular displacement, further analyses were conducted to investigate whether the differences in displacement could be related to the reconstruction parameters. Therefore, Kruskal–Wallis tests were performed with respect to the number of images, the number of triangles, and the texture resolution. The null hypothesis was rejected in all three cases, with p-value $= 0.004$ for the number of images, and p-value < 0.001 for both the number of triangles and texture resolution. Table 5 reports the pairwise comparisons, while Fig. 7 shows the distribution of the data among the reconstruction parameters. While the number of images and the number of triangles show only a limited number of significant differences, it is interesting to observe that, regarding texture, the lowest resolution (1 MP) exhibits significant differences in terms of total angular displacement with respect to all other resolution values.

5. Discussion

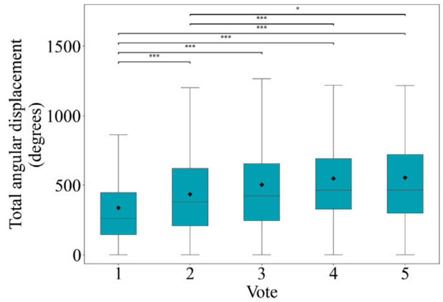
The goal of this work is to provide a comprehensive analysis of the various factors that may influence the perceived quality of reconstructed 3D models. The broader aim is to pave the way towards the definition of guidelines for the design of immersive experiences, involving 3D models, as well as the development of models suitable for scientific evaluations across diverse multimedia contexts, such as cultural heritage.

Regarding the first research question, **RQ1** (“How does perceived quality vary with reconstruction parameters such as the number of input images used, mesh complexity, and texture resolution?”), we analyzed the variation of MOS as a function of the reconstruction parameters. Results indicate that among the three factors – number of images, number of triangles, and texture resolution – the latter has the strongest influence on perceived quality. In particular, the MOS was significantly lower for 1 MP compared to higher resolutions (16 and 32 MP). The absence of substantial differences among 4, 16, and 32 MP suggests that, in multimedia contexts, texture resolution could be reduced (e.g., up to 16 MP) to optimize storage requirements without compromising perceived quality for end users. To better explain this consideration, Fig. 8 reports the MOS as a function of the model size. The gray covariance ellipses represent the joint dispersion of MOS and size in three size ranges (0 – 15 MB, 15 – 30 MB, over 30 MB). While some low-size models achieve high MOS values, they are more frequently associated with poor perceived quality, as evidenced by the presence of low MOS ratings. As the file size increases, the dispersion of MOS progressively decreases, leading to a consistently higher perceived quality. However, the MOS distributions associated with 16 MP and 32 MP textures largely overlap, suggesting a saturation effect in perceived quality. This behavior supports our claim that increasing texture resolution beyond 16 MP does not directly translate in higher perceived quality, while it significantly increases storage requirement. This makes intermediate texture resolutions a more efficient solution with respect to storage, without compromising visual perception.

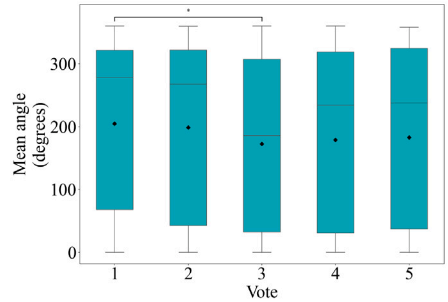
Concerning the number of images, the captures used for model reconstruction were uniformly distributed around the object to avoid visible errors that could have strongly affected the results. This acquisition strategy likely explains the absence of significant differences in perceived quality across different numbers of input images.

As to the number of triangles, statistical tests did not reveal significant pairwise differences, although some variations in data distribution were observed (Fig. 4(c)), especially between the lowest complexity level (5×10^3) and the two higher ones (50×10^3 and 100×10^3). This outcome may be partially attributed to the conservativeness of the Bonferroni correction – which, although stringent, ensures scientific robustness – but also to the rendering setup, which was performed on a desktop PC. In immersive scenarios where users can freely explore and perceive the full three-dimensional nature of the models, such as in VR, structural properties like geometric details are likely to play a more prominent role in influencing the perceived quality. In contrast, the desktop-based setup adopted in this study constrains user interaction to screen-based exploration, which may attenuate the perceptual relevance of geometric complexity and spatial depth. Therefore, while the present findings highlight the dominant role of texture resolution in desktop-based evaluations, they also suggest that future studies should explicitly investigate immersive settings, such as VR, to better capture the interplay between perceived quality, interaction behavior, and structural complexity. Such investigations are essential to generalize perceptual quality assessment frameworks to fully immersive multimedia experiences.

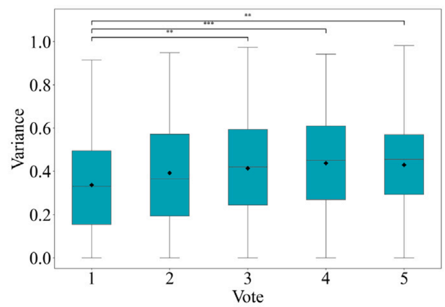
For the second research question, **RQ2** (“Are existing objective metrics correlated with human visual perception when applied to 3D reconstructions?”), image-based NR metrics demonstrated weak correlation with subjective ratings, whereas FR metrics showed a stronger



(a) Total angular displacement



(b) Mean angle



(c) Angle variance

Fig. 6. Data distribution for angular features with respect to the provided subjective scores. Outliers were removed for visualization purposes. The asterisks mark significance levels: '*' for $p < 0.05$, '**' for $p < 0.01$, '***' for $p < 0.001$.

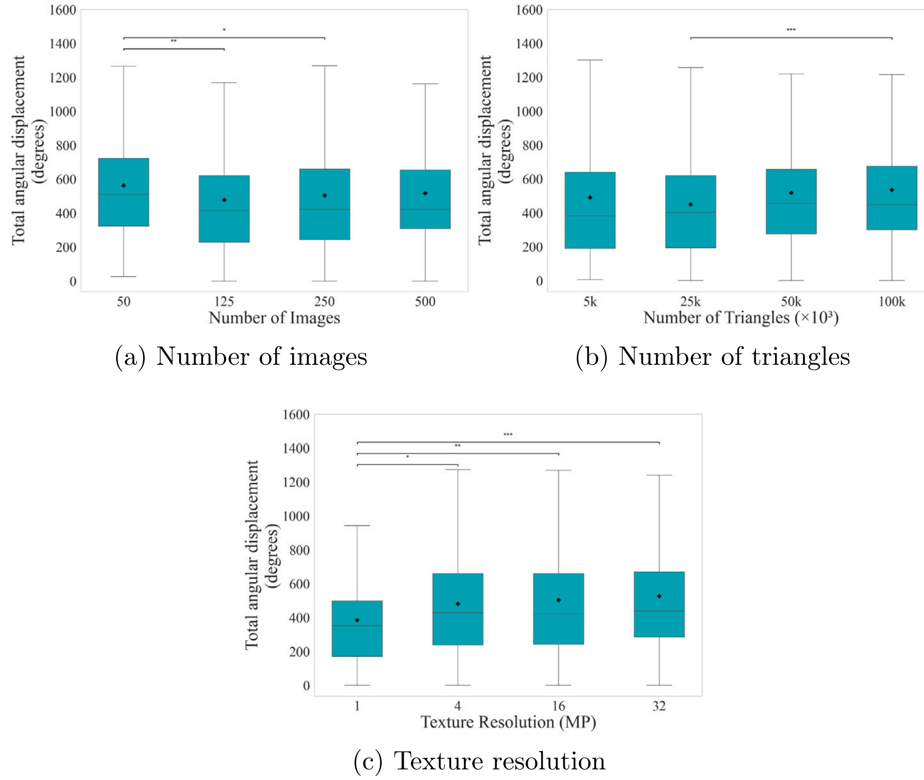
Table 5
Pairwise comparisons among reconstruction parameters for spatial exploration (total angular displacement).

Factor	Kruskal–Wallis p-value	Parameter comparison	Tukey HSD p-value
Number of Images	0.019*	50 vs. 125	0.002*
		50 vs. 250	0.045*
		50 vs. 500	0.264
		125 vs. 250	1.0
		125 vs. 500	1.0
		250 vs. 500	1.0
Number of triangles	<0.01*	5 vs. 25	1.0
		5 vs. 50	0.469
		5 vs. 100	0.077
		25 vs. 100	0.116
		25 vs. 50	<0.001*
		50 vs. 100	1.0
Texture resolution	<0.01*	1 vs. 4	0.024*
		1 vs. 16	0.002*
		1 vs. 32	<0.001*
		4 vs. 16	1.0
		4 vs. 32	1.0
		16 vs. 32	1.0

* Indicates a statistically significant result.

consistency with the scores provided by the participants. Among them, it is worth noting that PSNR, SSIM, and LPIPS correlated reasonably well with subjective perception. This behavior may be connected to the findings of **RQ1**, as these metrics capture pixel- or patch-level features on 2D projections of the model, which are closely related to the texture. As to FID, although not specifically designed for 3D reconstruction tasks, its distribution-based perceptual representation appears to capture global visual characteristics that are relevant to human perception. Although FR metrics are not directly applicable to real-life 3D reconstruction contexts, the outcome of the performed study paves the way for their applicability as benchmarks in the evaluation of novel NR metrics.

One of the most relevant outcomes concerns **RQ3** ("Is user interaction driven by model quality?"). These results can be discussed from two perspectives. The first relates to whether the interaction patterns are linked to subjective perception. Both total exploration time and total angular displacement showed significant differences across quality levels, particularly between the lowest and highest scores, while such differences tended to diminish among higher-quality models. The mean viewing angle, representing where users tended to focus, was highly variable regardless of the quality score, suggesting that it might depend more on the model content rather than on perceived quality. Although this aspect falls outside the scope of this paper, it represents an interesting direction for future research. The angular variance, representing the



(a) Number of images (b) Number of triangles

(c) Texture resolution

Fig. 7. Data distribution for the total angular displacement with respect to reconstruction parameters. Outliers were removed for visualization purposes. The asterisks mark significance levels: '*' for $p < 0.05$, '**' for $p < 0.01$, '***' for $p < 0.001$.

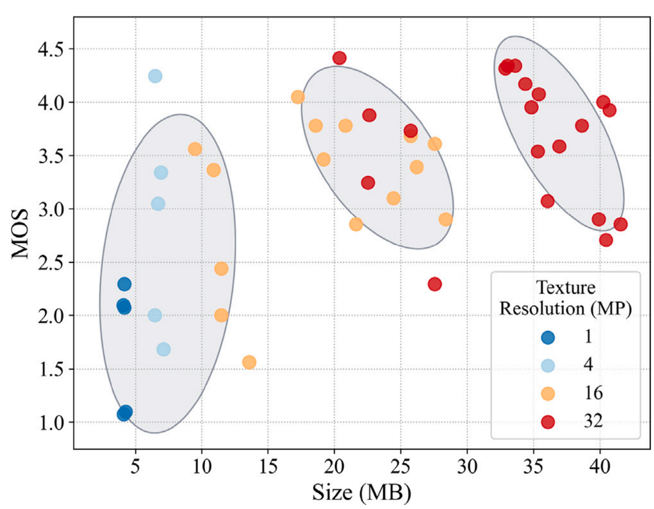


Fig. 8. MOS as a function of file size (MB). The gray areas represent covariance ellipses computed over the samples grouped by size ranges (0 – 15 MB, 15 – 30 MB, over 30 MB).

dispersion around the mean viewing direction, also showed differences — especially between ratings 1 and 3 or higher.

Overall, these results, particularly regarding total time and total angular displacement, highlight an important point: standard evaluation protocols, which often rely on video renderings of 3D models, fail to capture the effect of interaction on perceived quality. With the most novel technologies such as VR, perceived quality is no longer linked to only passive observation, but to an active immersive experience. Furthermore, the task associated with model exploration plays a crucial role. In the present study, participants likely interacted more with higher quality models because they were instructed to identify artifacts relative to the reference before assigning the rating. Under different tasks, interaction could instead mask or emphasize these artifacts — meaning that users might become more or less aware of distortions depending on how they engage with the content. Consequently, as immersive and non-immersive technologies evolve, there is a growing need to design experimental protocols that explicitly account for these aspects.

The second perspective concerns whether interaction behavior varies with reconstruction parameters. In this case, no significant differences were observed, except, once again, for texture resolution, which consistently emerged as the dominant factor.

6. Conclusions

This work presented a comprehensive analysis of how reconstruction parameters, objective metrics, and user interaction jointly shape the perceived quality of 3D reconstructions. The results indicate that texture resolution is the most influential factor on subjective quality, suggesting that perceptual optimization can be achieved without increasing geometric complexity or acquisition redundancy. The strong correlation between FR image-based metrics and subjective judgments confirms their suitability as perceptual indicators when reference data are available, which is not the case for 3D reconstruction. On the other hand, NR measures show inadequate performance in terms of their alignment with subjective perception, thus highlighting the need for new metrics. Interaction analysis revealed that users engage more with higher-quality models, both in terms of time and spatial exploration, emphasizing the need to integrate interactive behavior into quality assessment protocols — especially for immersive applications. Future research will focus on extending this framework to immersive settings, exploring perceptual effects under VR conditions, and analyzing how different 3D reconstruction modalities (e.g., NeRF, GS) influence perceived quality.

CRediT authorship contribution statement

Anna Ferrarotti: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Isabel Rodríguez:** Writing – original draft, Software, Conceptualization, Methodology. **Javier Usón:** Software, Conceptualization, Methodology, Writing – original draft. **Sara Baldoni:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **David Barbero:** Methodology, Software. **Daniel Berjón:** Supervision, Methodology, Conceptualization, Writing – review & editing. **Francisco Morán:** Supervision, Methodology, Writing – review & editing. **Narciso García:** Supervision, Methodology, Conceptualization, Writing – review & editing. **Federica Battisti:** Writing – review & editing, Supervision, Methodology, Conceptualization, Funding acquisition. **Jesús Gutiérrez:** Writing – review & editing, Supervision, Methodology, Conceptualization, Funding acquisition. **Marco Carli:** Writing – review & editing, Supervision, Methodology, Conceptualization, Funding acquisition. **Julián Cabrera:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Federica Battisti, given her role as editor-in-chief, had no involvement in the peer review of this article and did not have access to information regarding its peer review. Full responsibility for the editorial process for this artcle has been delegated to another journal editor.

Sara Baldoni, given her role as guest editor of the special issue “Immersive and networked media for next-gen communications”, had no involvement in the peer review of this article and did not have access to information regarding its peer review. Full responsibility for the editorial process for this article has been delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) Mission 4, Component 2, Investment 1.3, CUP C93C22005250001, partnership on “Telecommunications of the Future” (PE00000001 - program “RESTART”) and on the Spoke 4 - Cascade Call Project “EXperience and Privacy for Extended RealiTy” (“EXPERT”) and by the projects PID2023-148922OA-I00 (EEVOCATIONS) and TED2021-131690B-C31 (Revolution) funded by MCIN/AEI/10.13039/501100011033 of the Spanish Government, and HORIZON-IA-1010702-50 (XReco) funded by the European Union.

Data availability

Information about code and data is available at https://github.com/gti-upm/CH3D-Reco.

References

[1] E.H. Korkut, E. Surer, Visualization in virtual reality: a systematic review, Virtual Real. 27 (2) (2023) 1447–1480.
[2] T. Islam, A. Miron, X. Liu, Y. Li, Deep learning in virtual try-on: A comprehensive survey, IEEE Access 12 (2024) 29 475–29 502.
[3] K. Yu, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, D. Roth, Avatars for teleconsultation: Effects of avatar embodiment techniques on user perception in 3d asymmetric telepresence, IEEE Trans. Vis. Comput. Graphics 27 (11) (2021) 4129–4139.
[4] M. Milosz M. Skublewska-Paszkowska, P. Powroznik, E. Lukasik, 3D technologies for intangible cultural heritage preservation—literature review for selected databases, Herit. Sci. 10 (1) (2022) 3.
[5] C. Timmerer, S. Baraković, J. Baraković Husić, S. Bech, S. Bosse, J. Botev, K. Brunnström, L. Cruz, K. De Moor, A. de Polo Saibanti, et al., QUALINET white paper on definitions of immersive media experience (IMEx), 2020.
[6] P. Moulon, P. Monasse, R. Marlet, Global fusion of relative motions for robust, accurate and scalable structure from motion, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 3248–3255.
[7] B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng, NeRF: Representing scenes as neural radiance fields for view synthesis, Commun. ACM 65 (1) (2021) 99–106.
[8] B. Kerbl, G. Kopanas, T. Leimkuehler, G. Drettakis, 3D Gaussian splatting for real-time radiance field rendering, ACM Trans. Graph. 42 (4) (2023) [Online]. Available: http://dx.doi.org/10.1145/3592433.
[9] S. Doukianou, V. Lalioti, Ethical extended reality: Bridging technology and cultural heritage, 2024 IEEE International Conference on Metrology for EXtended Reality, Artificial Intelligence and Neural Engineering, MetroXRAINE, IEEE, 2024, pp. 301–306.
[10] S. Ubik, J. Kubišta, T. Dvořák, Interactive 3d models: Documenting and presenting restoration and use of heritage objects, Digit. Appl. Archaeol. Cult. Herit. 27 (2022) e00246.
[11] P. Singh, N. Pahuja, M. Kansal, S. Gurung, U. Shukla, S. Gupta, Enhancing tourism experiences and preserving cultural heritage with AR and VR, 2024 2nd International Conference on Disruptive Technologies, ICDT, IEEE, 2024, pp. 225–231.
[12] L. Cecere, F. Colace, M. De Santo, A. Lorusso, D. Santaniello, C. Valentino, Overview of cultural heritage education and emerging technologies, 2024 IEEE International Humanitarian Technologies Conference, IHTC, IEEE, 2024, pp. 1–7.
[13] L. Zhou, G. Wu, Y. Zuo, X. Chen, H. Hu, A comprehensive review of vision-based 3d reconstruction methods, Sens. 24 (7) (2024) 2314.
[14] J. Wu, O. Wyman, Y. Tang, D. Pasini, W. Wang, Multi-view 3d reconstruction based on deep learning: A survey and comparison of methods, Neurocomputing 582 (2024) 127553.
[15] S.Y. Bao, S. Savarese, Semantic structure from motion, CVPR 2011, IEEE, 2011, pp. 2025–2032.
[16] E. Prados, O. Faugeras, Shape from shading, Handbook of Mathematical Models in Computer Vision, Springer, 2006, pp. 375–388.
[17] P.-S. Tsai, M.A. Shah, Shape from shading with variable albedo, Opt. Eng., Bellingham 37 (4) (1998) 1212–1220.
[18] M. Belkaid, M. Merras, A. Berrajaa, N. El Akkad, et al., Review of 3D scene reconstruction: From traditional methods to advanced deep learning models, 2024 3rd International Conference on Embedded Systems and Artificial Intelligence, ESAI, IEEE, 2024, pp. 1–11.
[19] S. Liu, M. Yang, T. Xing, R. Yang, A survey of 3D reconstruction: The evolution from multi-view geometry to NeRF and 3DGS, Sens. 25 (18) (2025) 5748.
[20] C. Chen, Z. Xue, L. Yang, Z. Wu, S. Pang, H. Qin, A comprehensive survey on 3D single-view object reconstruction, IEEE Trans. Vis. Comput. Graphics 2025.
[21] A.J.H. Koh, S.Y. Tan, M.F. Nasrudin, A systematic literature review of generative adversarial networks (GANs) in 3D avatar reconstruction from 2D images, Multimedia Tools Appl. 83 (26) (2024) 68813–68853.
[22] M. Sajid, K.R. Malik, A.H. Khan, M. Fuzail, J. Li, Tvae-3d: Efficient multi-view 3d shape reconstruction with diffusion models and transformer based vae, Clust. Comput. 28 (13) (2025) 854.
[23] B.B. Bilecen, A. Gökmen, A. Dundar, Dual encoder GAN inversion for high-fidelity 3D head reconstruction from single images, Adv. Neural Inf. Process. Syst. 37 (2024) 87357–87385.
[24] A. Dalal, D. Hagen, K.G. Robbersmyr, K.M. Knausgård, Gaussian splatting: 3D reconstruction and novel view synthesis: A review, IEEE Access 12 (2024) 96797–96820.
[25] T. Installations, L. Line, ITU-T P. 910, in: Subjective Video Quality Assessment Methods for Multimedia Applications, Recommendation ITU-T, vol. 910, 2023.
[26] T. Installations, L. Line, ITU-T P. 919, in: Subjective Test Methodologies for 360° Video on Head-Mounted Displays, Recommendation ITU-T, vol. 919, 2020.
[27] J. Gutiérrez, G. Dandyyeva, M. Dal Magro, C. Cortés, M. Brizzi, M. Carli, F. Battisti, Subjective evaluation of dynamic point clouds: Impact of compression and exploration behavior, 2023 31st European Signal Processing Conference, EUSIPCO, IEEE, 2023, pp. 675–679.
[28] S. Rossi, I. Viola, P. Cesar, Behavioural analysis in a 6-DoF VR system: influence of content, quality and user disposition, in: Proceedings of the 1st Workshop on Interactive EXtended Reality, 2022, pp. 3–10.
[29] Y. Nehmé, J.-P. Farrugia, F. Dupont, P.L. Callet, G. Lavoué, Comparison of subjective methods for quality assessment of 3D graphics in virtual reality, ACM Trans. Appl. Percept. (TAP) 18 (1) (2020) 1–23.
[30] Y. Nehmé, J. Delanoy, F. Dupont, J.-P. Farrugia, P. Le Callet, G. Lavoué, Textured mesh quality assessment: Large-scale dataset and deep learning-based quality metric, ACM Trans. Graph. 42 (3) (2023) 1–20.
[31] D.V. Nguyen, T.T. Hien, T.T. Huong, A subjective quality evaluation of 3D mesh with dynamic level of detail in virtual reality, 2024 IEEE International Conference on Image Processing, ICIP, IEEE, 2024, pp. 1225–1231.

[32] P. Martin, A. Rodrigues, J. Ascenso, M.P. Queluz, Nerf view synthesis: Subjective quality assessment and objective metrics evaluation, IEEE Access 2024.
[33] S. Lee, J. Kang, S. Lee, W. Lin, A.C. Bovik, 3D-PSSIM: Projective structural similarity for 3D mesh quality assessment robust to topological irregularities, IEEE Trans. Pattern Anal. Mach. Intell. 46 (12) (2024) 9595–9611.
[34] Z. Zhang, Y. Zhou, W. Sun, X. Min, Y. Wu, G. Zhai, Perceptual quality assessment for digital human heads, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2023, pp. 1–5.
[35] Y. Nehmé, M. Abid, G. Lavoué, M.P. Da Silva, P. Le Callet, CMDM-VAC: Improving a perceptual quality metric for 3D graphics by integrating a visual attention complexity measure, 2021 IEEE International Conference on Image Processing, ICIP, IEEE, 2021, pp. 3368–3372.
[36] Z.C. Yildiz, A.C. Oztireli, T. Capin, A machine learning framework for full-reference 3d shape quality assessment, Vis. Comput. 36 (1) (2020) 127–139.
[37] X. Feng, W. Wan, R.Y. Da Xu, S. Perry, P. Li, S. Zhu, A novel spatial pooling method for 3D mesh quality assessment based on percentile weighting strategy, Comput. Graph. 74 (2018) 12–22.
[38] Z. Zhang, Y. Zhou, C. Li, K. Fu, W. Sun, X. Liu, X. Min, G. Zhai, A reduced-reference quality assessment metric for textured mesh digital humans, ICASSP, IEEE, 2024, pp. 2965–2969.
[39] I. Abouelaziz, M. Omari, M.E. Hassouni, H. Cherifi, Reduced reference 3d mesh quality assessment based on statistical models, in: 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems, SITIS, 2015, pp. 170–177.
[40] I. Abouelaziz, A. Chetouani, M. El Hassouni, L.J. Latecki, H. Cherifi, No-reference mesh visual quality assessment via ensemble of convolutional neural networks and compact multi-linear pooling, Pattern Recognit. 100 (2020) 107174.
[41] 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.
[42] Z. Zhang, W. Sun, Y. Zhou, W. Lu, Y. Zhu, X. Min, G. Zhai, EEP-3DQA: Efficient and effective projection-based 3D model quality assessment, 2023 IEEE International Conference on Multimedia and Expo, ICME, IEEE, 2023, pp. 2483–2488.
[43] Z. Zhang, W. Sun, H. Wu, Y. Zhou, C. Li, Z. Chen, X. Min, G. Zhai, W. Lin, GMS-3DQA: Projection-based grid mini-patch sampling for 3D model quality assessment, ACM Trans. Multimed. Comput. Commun. Appl. 20 (6) (2024) 1–19.
[44] M. El Hassouni, H. Cherifi, Learning graph features for colored mesh visual quality assessment, 2022 IEEE International Conference on Image Processing, ICIP, IEEE, 2022, pp. 3381–3385.
[45] G. Li, W. Gao, W. Gao, MPEG video-based point cloud compression (v-PCC) standard, Point Cloud Compression: Technologies and Standardization, Springer, 2024, pp. 199–218.
[46] 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. 26 (2024) 6730–6742.
[47] A. Ferrarotti, I. Rodríguez, J. Usón, S. Baldoni, J. Gutiérrez, D. Berjón, F. Morán, F. Battisti, N. García, M. Carli, J. Cabrera, Analysis of objective 3D mesh quality metrics for cultural heritage, 2025 15th International Conference on Quality of Multimedia Experience, QoMEX, IEEE, 2024.
[48] P. Moulon, P. Monasse, R. Perrot, R. Marlet, OpenMVG: Open multiple view geometry, International Workshop on Reproducible Research in Pattern Recognition, Springer, 2016, pp. 60–74.
[49] OpenMVS: Open multi-view stereo reconstruction library, 2015, https://cdcseacave.github.io/. (Accessed 03 April 2025).
[50] M. Kazhdan, M. Bolitho, H. Hoppe, Poisson surface reconstruction, in: Proceedings of the fourth Eurographics symposium on Geometry processing, vol. 7, p. 2006, no. 4.
[51] R. Pagés, D. Berjón, F. Morán, N. García, Seamless, static multi-texturing of 3D meshes, Comput. Graph. Forum 34 (1) (2015) 228–238.
[52] European Parliament and Council of the European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council. [Online]. Available: https://data.europa.eu/eli/reg/2016/679/oj.
[53] S.A. Nolan, T.E. Heinzen, Statistics for the Behavioral Sciences, Macmillan, 2011.
[54] G. Marsaglia, W.W. Tsang, J. Wang, Evaluating Kolmogorov's distribution, J. Stat. Softw. 8 (2003) 1–4.
[55] M.B. Brown, A.B. Forsythe, Robust tests for the equality of variances, J. Amer. Statist. Assoc. 69 (346) (1974) 364–367.
[56] J.W. Tukey, Comparing individual means in the analysis of variance, Biometrics (1949) 99–114.
[57] W.H. Kruskal, W.A. Wallis, Use of ranks in one-criterion variance analysis, J. Amer. Statist. Assoc. 47 (260) (1952) 583–621.
[58] H.B. Mann, D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. (1947) 50–60.
[59] A. Guédon, V. Lepetit, SuGaR: Surface-aligned Gaussian splatting for efficient 3D mesh reconstruction and high-quality mesh rendering, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 5354–5363.
[60] H. Jiang, Z. Jiang, Y. Zhao, Q. Huang, Leap: Liberate sparse-view 3d modeling from camera poses, 2023, arXiv preprint arXiv:2310.01410.
[61] X. Lyu, Y.-T. Sun, Y.-H. Huang, X. Wu, Z. Yang, Y. Chen, J. Pang, X. Qi, 3dgsr: Implicit surface reconstruction with 3d Gaussian splatting, ACM Trans. Graph. 43 (6) (2024) 1–12.
[62] A. Obukhov, M. Krasnyanskiy, Quality assessment method for GAN based on modified metrics inception score and Fréchet inception distance, in: Proceedings of the Computational Methods in Systems and Software, Springer, 2020, pp. 102–114.