Visual Quality of 3D Meshes with Diffuse Colors in Virtual Reality: Subjective and Objective Evaluation

Yana Nehmé, Florent Dupont, Jean-Philippe Farrugia, Patrick Le Callet, *Fellow, IEEE*
and Guillaume Lavoué, *Senior Member, IEEE*

Abstract—Surface meshes associated with diffuse texture or color attributes are becoming popular multimedia contents. They provide a high degree of realism and allow six degrees of freedom (6DoF) interactions in immersive virtual reality environments. Just like other types of multimedia, 3D meshes are subject to a wide range of processing, e.g., simplification and compression, which result in a loss of quality of the final rendered scene. Thus, both subjective studies and objective metrics are needed to understand and predict this visual loss. In this work, we introduce a large dataset of 480 animated meshes with diffuse color information, and associated with perceived quality judgments. The stimuli were generated from 5 source models subjected to geometry and color distortions. Each stimulus was associated with 6 hypothetical rendering trajectories (HRTs): combinations of 3 viewpoints and 2 animations. A total of 11520 quality judgments (24 per stimulus) were acquired in a subjective experiment conducted in virtual reality. The results allowed us to explore the influence of source models, animations and viewpoints on both the quality scores and their confidence intervals. Based on these findings, we propose the first metric for quality assessment of 3D meshes with diffuse colors, which works entirely on the mesh domain. This metric incorporates perceptually-relevant curvature-based and color-based features. We evaluate its performance, as well as a number of Image Quality Metrics (IQMs), on two datasets: ours and a dataset of distorted textured meshes. Our metric demonstrates good results and a better stability than IQMs. Finally, we investigated how the knowledge of the viewpoint (i.e., the visible parts of the 3D model) may improve the results of objective metrics.

Index Terms—Computer Graphics, Perception, Virtual reality, Diffuse Color, 3D Mesh, Visual Quality Assessment, Subjective Quality Evaluation, Objective Quality Evaluation, Dataset, Perceptual Metric.

1 Introduction

As technological advances and capabilities in the field of computer graphics grow day by day, the need to master the manipulation, visualization and processing of 3D digital data increases at an equal pace. Indeed, the development of modeling software and acquisition devices (3D scan, reconstruction process) makes 3D graphics (mesh, voxel, point cloud) rich and realistic: complex models with millions of geometric primitives, enriched with various appearance attributes (color, texture, materiall, etc.). The way in which this 3D content is consumed is also evolving from standard screens to Virtual and Mixed Reality (VR/MR). However, the size and complexity of these rich 3D models often make their interactive visualization problematic. This is particularly the case in immersive environments (using head-mounted displays) and/or in case of online applications (where fast transmission is needed). Thus, to adapt the complexity of the 3D content for lightweight devices and to avoid latency due to transmission, diverse processing operations, including simplification and compression, are usually applied. These processes are lossy. They operate on both geometry and appearance attributes, which inevitably introduce distortions that impact the perceived quality of the data and thus the quality of user experience (QoE).

Objective quality metrics are thus critically needed to automatically predict the level of annoyance caused by these operations. Most metrics in the literature evaluate only geometric distortions (i.e. they consider meshes without appearance attributes), e.g. [1], [2], [3]. When it comes to meshes with diffuse color information (either in the form of texture or vertex-colors), little work has been published [4] [5]. Actually, for this kind of data, it is still unclear how color and geometry distortions affect quality. There is a lack of both objective metrics and subjective datasets. Another factor that has not yet been explored, and which is relevant in the case of 6 Degrees of Freedom (DoF) interactions, is how the viewpoint and movement of 3D models affect their perceived quality.

In this work, we address the problem of subjective and objective quality assessment of 3D models with diffuse colors. Our first goal is to produce a ground truth database of 3D graphics with quality judgments, and to understand the impact of several factors (such as the source models, distortions, viewpoints, and animations) on the perceived quality of these data. The experiment is based on a double stimulus impairment scale (DSIS) method and involves 480 animated stimuli created from five source models, each corrupted with color and geometry distortions and displayed in 3 different viewpoints that we animated with 2 short movements. We chose to conduct the experiment in Virtual

- *Y. Nehmé, F. Dupont, JP. Farrugia and G. Lavoué are with the CNRS, Univ Lyon, LIRIS, France.*
  E-mail: yana.nehme@insa-lyon.fr, jean-philippe.farrugia@univ-lyon1.fr, Florent.Dupont@liris.cnrs.fr, glavoue@liris.cnrs.fr.
- *P. Le Callet is with the CNRS, Univ Nantes, LS2N, France.*
  E-mail: patrick.lecallet@univ-nantes.fr.

Manuscript submitted June 16, 2020.

Reality (VR) using the HTC Vive Pro headset, since VR user studies offer the most ecological and realistic use cases and are in high demand. This database is used to analyze the factors that influence the subjective quality assessment of 3D graphics: we evaluate not only the visual impact of color and geometry distortions on the appearance of such data, but also the impact of source models, animations and viewpoints.

Considering the findings of this subjective evaluation, we design an objective quality assessment metric for colored meshes: *CMDM* (Color Mesh Distortion Measure). This is a full-reference data-driven metric that fully operates on the mesh domain, at vertex level. It consists of a linear combination of perceptually-relevant features related to color and geometry. The optimal set of features was selected through logistic regression. We use our subjective ground truth to evaluate the performance of *CMDM* as well as other state-of-the-art image quality metrics. Moreover, to assess the robustness of our metric, we test it on a new dataset of textured meshes corrupted with compound distortions that differ significantly from those used to train it. Our metric demonstrates good results. Finally, we study the relevance of incorporating the viewpoint (the visible parts) of the 3D model into objective metrics.

We summarize our contributions as follows:

1) We provide the community with a ground truth dataset of 480 animated meshes with vertex colors, each rated by 24 subjects. This dataset is the largest one for quality assessment of 3D contents with diffuse color information, and the first based on vertex color representation. It is also the first public dataset[1] produced in VR for such data.
2) We provide an in-depth analysis of the effects of source models, distortions, viewpoints and movement on both mean opinion scores (MOSs) and confidence intervals (CIs). Our findings provide insights for the design of both subjective studies and objective metrics.
3) We evaluate individually the performance of a set of perceptually-relevant curvature-based and color-based features for predicting the perceived visual quality of colored meshes.
4) We develop and learn a perceptually-validated metric for measuring the quality of colored meshes. To the best of our knowledge, our proposed metric is the first attempt to integrate both geometry and color information for quality assessment of such data. Our metric demonstrates good results and stability on two different datasets. The source code of the metric is made publicly available[2] on the MEsh Processing Platform (MEPP) to support further research in this area.
5) We investigate how knowledge of the viewpoint may improve results from objective metrics.

The paper is organized as follows: section 2 provides a review of the existing works on subjective and objective quality assessment of 3D data. In section 3, we describe the subjective study, before presenting the results in section 4. Section 5 details the proposed metric, while section 6 presents its validation as well as a comparison with state-of-the-art image and mesh quality metrics. The study on integrating the viewpoint in objective metrics is presented in section 7 along with its results. Finally, concluding remarks and perspectives are outlined in section 8.

2 Related work

In this section, we provide an overview of existing datasets and metrics for predicting the perceived visual impact of distortions applied to graphical 3D content (3D meshes and point clouds). We are specifically interested in 3D content with diffuse colors, either in the form of texture maps or vertex/point colors. Note also that this state-of-the-art focuses on the visual impact of distortions applied on the 3D content itself (e.g. introduced by compression, simplification or filtering); it does not cover the visibility prediction of artifacts introduced during the rendering process (e.g. by global illumination approximation) or after rendering (e.g. by tone mapping). A dataset has been recently introduced that focuses on these types of artifacts [6]. For a more complete survey of the field of perception and quality assessment in computer graphics, we refer the reader to [7].

2.1 Subjective quality experiments and datasets

As stated in the introduction, datasets of human perceptual similarity judgments are of primary importance for understanding human behavior in evaluating perceived quality, as well as for training and benchmarking objective metrics. Many authors have conducted subjective quality assessment tests involving 3D meshes [2], [5], [8], [9], [10], [11], [12], [13], [14], [15], [16] or 3D point clouds [17], [18], [19], [20], [21]. They considered a variety of methods: Absolute Category Rating (ACR) [11], [13], [22], Double-Stimulus Impairment Scale (DSIS) [2], [8], [9], [18], [19], [21] and Pairwise Comparison (PC) [5], [14], [15], [23]. Very recently, a study [16] attempted to compare these subjective methods and showed that, for the particular case of 3D graphical content, the DSIS method tends to produce more accurate results than ACR (i.e., MOS with smaller confidence intervals). Existing subjective experiments also considered different ways of presenting 3D content: static images [8], animated content without interaction (usually low-speed rotation) [5], [9], [10], [16], [19], [20], [21], [22], [23] or interactive content [2], [11], [13], [14], [15], [17], [18]. We denote that only in [15], [16], the experiments were conducted in a VR environment. So far, no attempts have been made to fully understand the impact of these design choices on the obtained mean opinion scores and their accuracy.

Unfortunately, among the works presented above, very few have publicly released their datasets. For 3D meshes, the available datasets of mean opinion scores concern mostly geometry-only content [11], [12], [14] and are all rather small (resp. 88, 26 and 30 models). The only public datasets involving 3D meshes with diffuse color information are provided by Guo et al. [5] and from Zerman et al. [22], and contain respectively 136 and 28 stimuli. For both cases, the color information is provided as texture maps. For colored point clouds, the available datasets are those provided by Javaheri et al. [20], Alexiou et al. [18] and Zerman et al. [22], and contain respectively 54, 244 and 136 stimuli. Note that

[1] https://yananehme.github.io/
[2] https://github.com/MEPP-team/MEPP2

the dataset from Zerman et al. [22] actually contains both meshes and point clouds, for a total of 164 stimuli that were rated in the same subjective test. All these datasets were generated through experiments conducted on screen.

In this work, we propose a dataset of 480 animated meshes with vertex colors. It is the largest one for quality assessment of 3D content with diffuse color information, and the first based on vertex color representation. Note that the *mesh* representation differs considerably from the *point cloud* representation in several aspects such as the way they are rendered, and the nature of commonly applied processing operations (and thus distortions). Our dataset allows us to provide an initial investigation on the influence of movement and viewpoint on the quality evaluation of 3D content. As in [15], [16] we considered a VR context and we used a DSIS method with 24 observers per stimulus, as recommended in [16].

2.2 Objective quality metrics

Inspired by the vast amount of previous works on image and video quality assessment, several objective quality metrics have been introduced for 3D meshes. These are mostly full-reference (compare the distorted model with its original/reference) and follow the classical approach used in image quality assessment: local feature differences are calculated at vertex level, which are then pooled over the entire 3D model to obtain a global quality score. Existing metrics rely on various geometry characteristics: curvature [1], [24], dihedral angles [2] or roughness [3], [13]. A survey [25] showed that MSDM2 [1], FMPD [3] and DAME [2] are excellent predictors of visual quality. Very recently, several authors proposed data-driven approaches based on machine learning [26], [27]. Besides these works on global visual quality assessment (suited for supra-threshold distortions), Nader et al. [28] introduced a bottom-up visibility threshold predictor for 3D meshes. Guo et al. [29] also studied the local visibility of geometric artifacts and showed that curvature could be a good predictor of distortion visibility.

The above-presented works only take geometry into account. With respect to 3D content with color or material information, very few works have been published. For meshes with diffuse texture, Tian et al. [4] and Guo et al. [5] proposed metrics based on a weighted combination of a global distance over geometry (Mean Squared Error (MSE) over mesh vertices in [4], and MSDM2 in [5]) and a global distance over texture image (MSE over texture pixels in [4], and SSIM in [5]). While the latter metric demonstrated good results on a subjective dataset of distorted textured meshes [5], combining errors computed on different domains (3D mesh and texture image) may be hazardous since many external factors (e.g. texel size, visibility of different parts) may impact the results.

With regard to this previous work [4], [5], we propose a data-driven metric that fully operates on the mesh domain, at vertex level. We consider an initial collection of perceptually-relevant features related to color and geometry. These features are taken from existing works on quality assessment of 3D meshes [1] and color images [30]. A subset of these features is then optimally selected and combined, based on the results of our subjective study. Our metric provides excellent results and demonstrates a good stability, both for meshes with vertex colors and for textured meshes.

For quality assessment of 3D colored point clouds, a data-driven metric (PCQM) has been recently introduced by Meynet et al. [31]. Our metric considers the same initial collection of color and geometric features as [31]. Nevertheless, moving from point cloud domain to mesh domain implies major adaptations in the computation of these features. Other differences between our metric and PCQM are: the optimal selection and combination of features, the multi-scale approach, and the viewpoint integration mechanism. Our metric is also related to the work of Vanhoey et al. [23], who proposed a quality metric for surface light-fields (i.e., per-vertex directional color). However, their metric considers color information only and is actually a simple MSE over both the directional and spatial domains.

All the metrics presented above are model-based, i.e., they operate on the 3D model itself (or its attributes like texture maps). However, to evaluate the quality of 3D content, several authors have also considered Image Quality Metrics (IQM) computed on rendered snapshots. For example, Yang et al. [32] and Caillaud et al. [33] respectively used image MSE and SSIM [34] to optimize textured mesh transmission. The advantage of image-based metrics over model-based metrics is their natural ability to handle the multimodal nature of data (geometry and color or texture information), as well as their natural incorporation of the complex rendering pipeline (computation of light material interactions, viewpoint selection and rasterization). On the other hand, IQMs pose other problems: (1) it is necessary to know in advance the final rendering of the stimuli in order to predict their quality with these metrics (because IQMs operate on 2D rendered snapshots). (2) IQMs also need the knowledge of the displayed viewpoint. Using them in a view-independent way introduces new parameters such as choice of the 2D views, or pooling of quality scores obtained from different views into a single global score. (3) IQMs are not practical for driving processing operations (e.g. mesh simplification). Model-based metrics are better suited for these operations since they operate on the mesh domain, i.e. the same representation space as mesh processing algorithms. This makes it possible to drive a process globally (on the entire mesh) as well as locally (at vertex level). (4) Recent studies about these view-based approaches [5], [35] tend to show that their performance greatly depends on distortions and contents, and fall well behind model-based approaches.

3 Subjective experiment

We conducted a large-scale subjective experiment to evaluate the visual impact of color and geometry distortions on the appearance of colored 3D models. Our dataset contains 480 animated 3D models created from five reference objects, on which are applied four types of distortion and two types of animation. This dataset extends the one presented in [16] composed of 80 stimuli. The subjective study was conducted in a virtual reality setting using a DSIS method. This section provides details on the subjective study.

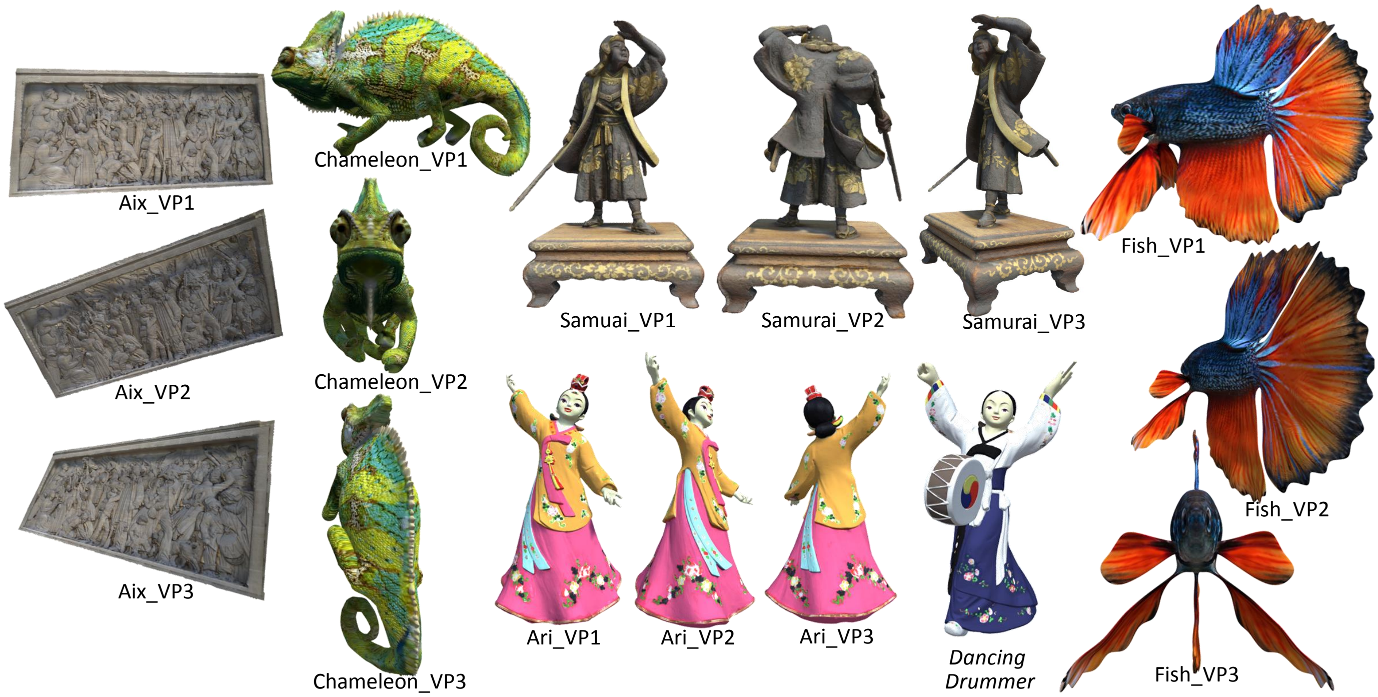


Fig. 1: Illustration of the 3D graphic source models and their selected viewpoints, respectively. Acronyms refer to Model_Viewpoint.

3.1 Stimuli

3.1.1 3D source model selection

To build our dataset of colored 3D models, we selected 5 high-resolution triangle meshes, each having diffuse color information represented by vertex colors (no texture mapping). These models were chosen so as to ensure a variety of shapes and colors. Table 1 details the characteristics of the models, while Figure 1 illustrates them. Note that, the sixth model (*"Dancing Drummer"*) is not part of the dataset. It was used at the training stage of the experiment.

TABLE 1: Characteristics of the 3D graphic source models

Models	**#Vertices**	**Geometry complexity**	**Color characteristics**	**Semantic category**	**Created using**
Aix	686061	Plane with small details	Mono-color	Art	3D scanning
Ari	645492	Intermediate	Cool & light colors	Human statues	3D scanning
Chameleon	588441	High & sharp edges	Cool & dull colors	Animal	Modeling software
Fish	216578	Low & sharp edges	Cool & warm colors	Animal	Modeling software
Samurai	449997	High	warm colors	Human statues	3D scanning
Dancing Drummer	1335436	Intermediate/ High	Cool colors	Human statues	3D scanning

3.1.2 Distortions

The source models presented above have been corrupted by the following 4 types of distortion applied on geometry and color. These selected distortions represent common simplification and compression operations typically used in 3D model modeling and post-processing. They are described below.

- Uniform geometric quantization (QGeo): applied on geometry. This is a very common process for lossy compression.
- Uniform LAB color quantization (QCol): applied on vertex colors. This is inspired by the usual 2D image compression processes.
- "Color-ignorant" simplification (SGeo): a surface simplification algorithm that takes into account geometry only. It consists of iterative edge collapse operations driven by the quadric error metrics [36].
- "Color-aware" simplification (SCol): a surface simplification algorithm that takes into account both geometry and color. It consists of iterative vertex removal operations, driven by a combination of (1) a geometry metric: the area loss caused by the removal; and (2) a color metric: the LAB distance between the color of the vertex to be removed and its interpolation after removal [37].

Each distortion was applied with 4 different strengths, adjusted manually in order to span the whole range of visual quality from imperceptible levels to high levels of impairment (as is typically the case in subjective image quality studies [38]). Figure 2 illustrates some visual examples, while all details about the distortion parameters are available in Table 2.

Thus, we generated 80 distorted models (5 source models × 4 distortion types × 4 strengths).

3.1.3 Stimuli generation

In existing subjective studies involving 3D content, different methods have been used to display the 3D models to the observers: still images, free interaction or animations. As shown by Rogowitz et al. [9], still images are not sufficient to evaluate the visual quality of 3D models. Thus, it is important that the object moves so that the observer can see the dynamic effects of shading on the shape. It is also important for the observer to see the whole object and not to focus on one single viewpoint. However, allowing free interaction leads to a cognitive overload which may alter human judgments. Inspired by the principle of pseudo-videos, 3D animation is a simple yet efficient way to control

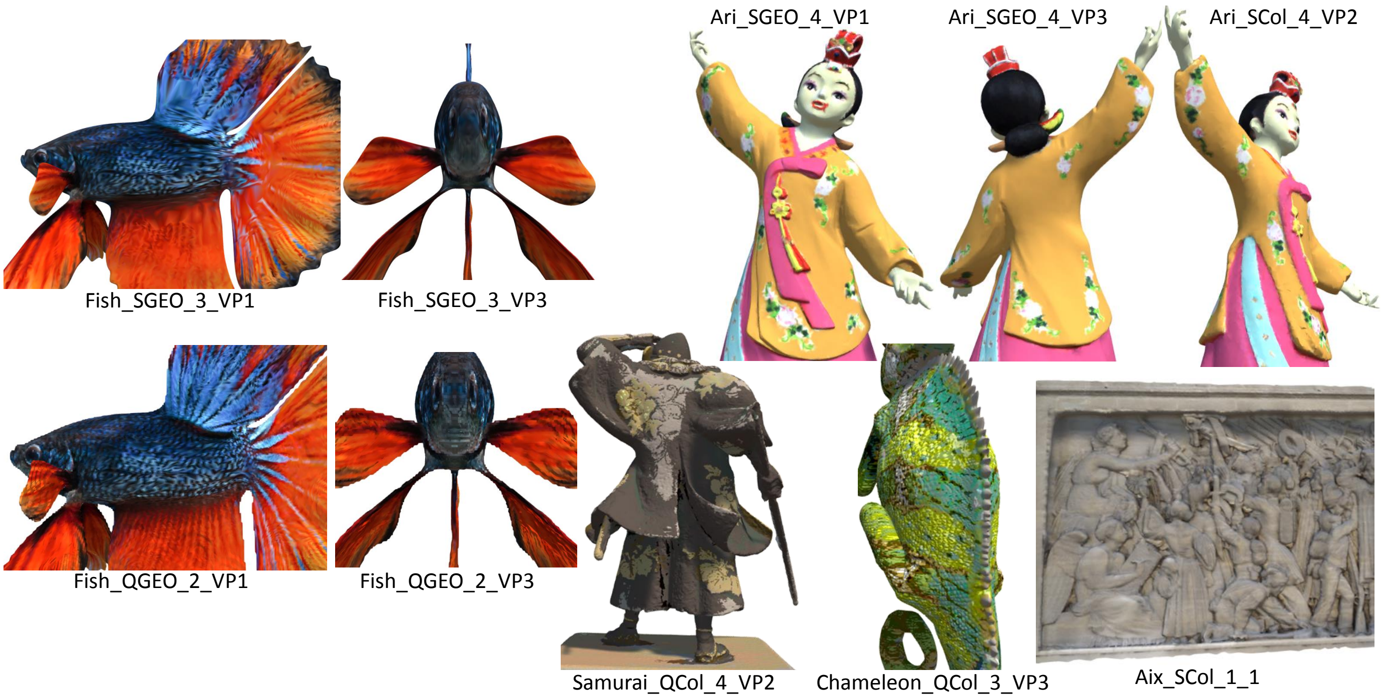


Fig. 2: Some examples of distorted models. Acronyms refer to Model_Dist-Type_Dist-Strength_Viewpoint.

TABLE 2: Details on the distortions applied to each source model.

Distortion type	Distortion strength	Aix	Ari	Chameleon	Fish	Samurai
QGeo	1	10 bits	10 bits	9 bits	9 bits	10 bits
	2	9 bits	9 bits	8 bits	8 bits	9 bits
	3	8 bits	8 bits	7 bits	7 bits	8 bits
	4	7 bits	7 bits	6 bits	6 bits	7 bits
QCol	1	(L=5, A=4, B=4) bits	(L=5, A=4, B=4) bits	(L=4, A=3, B=3) bits	(L=5, A=5, B=5) bits	(L=4, A=3, B=3) bits
	2	(L=4, A=3, B=3) bits	(L=4, A=3, B=3) bits	(L=3, A=2, B=2) bits	(L=4, A=3, B=3) bits	(L=4, A=2, B=2) bits
	3	(L=3, A=2, B=2) bits	(L=2, A=3, B=3) bits	(L=2, A=2, B=2) bits	(L=3, A=2, B=2) bits	(L=3, A=2, B=2) bits
	4	(L=2, A=2, B=2) bits	(L=3, A=3, B=3) bits	(L=2, A=1, B=1) bits	(L=2, A=2, B=2) bits	(L=2, A=2, B=2) bits
SGeo	1	50% removed	30% removed	50% removed	31% removed	24% removed
	2	75% removed	50% removed	75% removed	50% removed	50% removed
	3	88% removed	75% removed	87% removed	77% removed	75% removed
	4	94% removed	87% removed	92% removed	88% removed	88% removed
SCol	1	71% removed	50% removed	67% removed	77% removed	66% removed
	2	87% removed	64% removed	83% removed	79% removed	80% removed
	3	94% removed	88% removed	92% removed	87% removed	90% removed
	4	98% removed	94% removed	95% removed	96% removed	96% removed

the interaction between subject and stimulus. So as a compromise, we selected, for each model, 3 viewpoints that we animated with 2 short movements. These 6 combinations of viewpoints and movements can be considered to be the hypothetical rendering trajectories (HRTs), concept introduced in [39] for free-viewpoint videos. HRTs reflect/represent the dimension of the test object related to the interactivity part such as the camera configurations, viewpoints, trajectories of 3D objects.

In our experiment, the viewpoints were perceptually chosen and adjusted by experts, so that *viewpoint 1* represents the one which contains the most geometry, color and semantic information. *Viewpoint 2* and *viewpoint 3* cover the remaining semantically relevant parts of the model (see Figure 1). For each viewpoint of a given stimulus, we applied 2 types of animation:

- Slow rotation (R) of 15 degrees around the vertical axis in a clockwise and then in a counterclockwise direction.
- Slow zoom in (Z) of 0.75 meters, followed by a zoom out.

Note that, the animations we generate do not involve non-rigid transformations of the objects. Generating different stimulus orientations and animations will allow us to explore the impact of animations and viewpoints (HRT) on the perceived quality.

Our dataset thus contains 480 dynamic stimuli (5 source models × 4 distortion types × 4 strengths × 3 viewpoints × 2 animation types).

3.2 Experimental procedure and apparatus

The objective of our experiment is to produce a ground truth of subjective opinions on our set of 480 stimuli. We selected the Double Stimulus Impairment Scale (DSIS) methodology, as the subjective rating method. The observer sees the reference model and the same model impaired, simultaneously, side by side, for 10s and rates the impairment of the second stimulus in relation to the reference using a five-level impairment scale [40]. This method was shown to be more stable and more accurate than the Absolute Category Rating with Hidden Reference (ACR-HR) method for assessing the quality of 3D models [16]. Indeed, the authors showed that the presence of an explicit reference greatly improves the accuracy of results and reduces confidence intervals. This is due to the fact that people have less prior knowledge on 3D graphic quality compared to natural images/videos.

We chose to conduct the experiment in a fully immersive virtual environment (VE) since Virtual Reality is becoming a popular way of consuming and visualizing 3D content. We used the HTC Vive Pro headset, a high-end virtual reality headset with a resolution of 1440 x 1600 pixels per eye (2880 x 1600 pixels combined), a field of view of 110 degrees and a refresh rate of 90 Hz. The reference and the distorted model were rendered in a virtual scene, side by side, at a viewing distance fixed at 3 meters from the observer, under a given viewpoint and type of animation. Note that these 2 dynamic stimuli were specifically oriented in order to show exactly the same vertices of the 2 models at the same time. Their size is approximately 37 degrees of visual angle. Their material type complies with the Lambertian reflectance model (diffuse surfaces). The apparent brightness of such a surface to an observer is the same regardless of the observer′s angle of view/position in the scene. The stimuli are visualized in a neutral virtual room (light gray walls) without shadows and under a directional light (all the vertices are illuminated as though the light were always from the same direction.). We aimed to design a neutral room so that the experimental environment does not influence the users' perception of the stimulus.

We integrated the rating billboard in the VE of our experiment. This board is displayed after the presentation of each pair of stimuli. There is no time limit to vote and the stimuli are not shown during that time. The same neutral room (light gray walls), utilized to show the stimuli, is used in the rating environment. To vote, the subject selects and saves the score using the trigger of the HTC Vive controller. As in [41], to facilitate the interaction with the rating panel, we attached a laser beam to the controller. Figure 3 illustrates the experimental environment.
The entire experiment was developed with Unity3D using c# scripting.

3.3 Participants and training

Training: As recommended in the ITU-R BT.500 [42], the experiment started with training, during which observers familiarized themselves with the virtual environment and the task. We selected a training 3D model not included in our original test set: ”Dancing Drummer” (see Figure 1) and generated 11 distorted models that span the whole range of distortions. After each stimulus (displayed for 10s), the rating panel is displayed for 5s. An example score assigned to this distortion is highlighted. We added a practice trial stage at the end of the training: we displayed 2 extra stimuli and asked the subject to rate the quality or the impairment. The results of these stimuli were not recorded. This stage was used to allow the observer to familiarize him/herself with the experimentation, to focus appropriately, and to ensure that observers fully understand the task of the experiment.

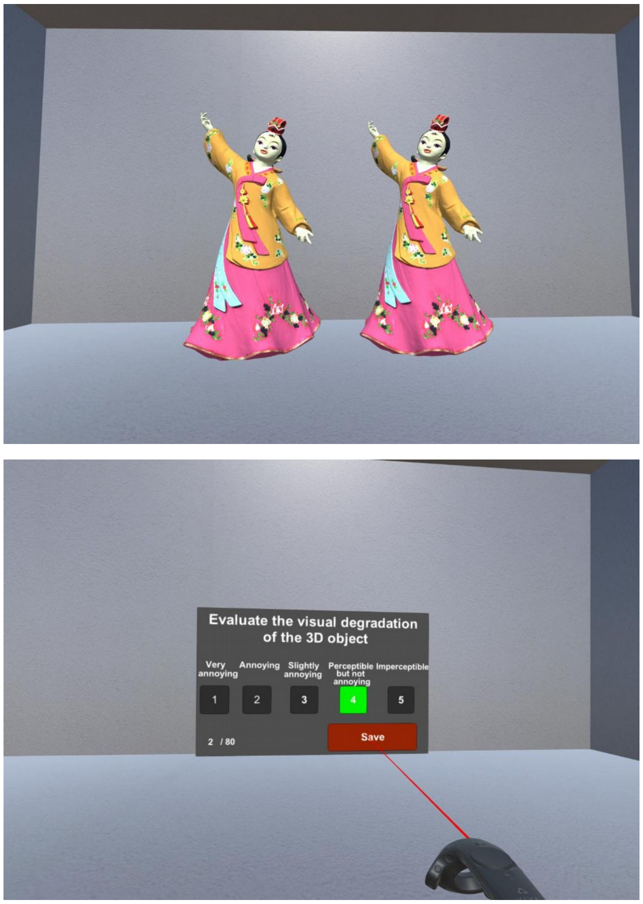

Evaluate the visual degradation of the 3D object
Very annoying
Annoying
Slightly annoying
Perceptible but not annoying
Imperceptible
1
2
3
4
5
2 / 80
Save

Fig. 3: The experimental environment of our subjective test based on the DSIS method.

Creation of test sessions: In order to maintain a sufficient level of attention, we decided to limit the number of stimuli rated per participant to 160 stimuli out of 480. As we specifically aim to assess whether the animation influences the ratings, we decided to show each participant one viewpoint in both rotation and zoom animations. Furthermore, we wanted each participant to see all the reference models, where each model is corrupted by all the distortion types and levels.
According to the recommendations of [16] about the required number of observers for assessing the quality of 3D data using the DSIS method, each stimulus must be rated by at least 24 observers.
With this in mind, we have developed an algorithm which creates, for each subject, a random batch of 160 stimuli respecting 2 constraints: (1) each batch must contain 5 reference models x 4 distortion types x 4 strengths x 1 viewpoint x 2 animations. (2) each stimulus must be rated 24 times (i.e. present in 24 batches).

Participants: A total of 72 (480*24/160) subjects took part in the experiment and they were remunerated. Participants were aged between 18 and 55. The majority were students from the University of Nantes, University of Lyon and LIRIS laboratory, while the rest were workers and professionals in different occupations. 49 males and 23 females, 45 of whom had already tried (or were familiar with) a VR headset, they were naive about the purpose of the experiments. All observers had a normal or corrected to normal vision.

Duration: To avoid fatigue, boredom and cyber sickness, we divided the 160 stimuli into 2 sessions of 23 min each (informed consent/instructions + 11 training stimuli x (10s display + 5s rating) + 80 test stimuli x (10s display + ≈4s rating)). None of these sessions took place on the same day in order to prevent any learning effect between stimuli. Thus, these two sessions were held at least two days apart. The stimuli were displayed in a random order (3D models, distortion types, levels and animations all mixed) to each observer. Each stimulus was presented once; the observer was not able to replay the objects.

4 Analysis of subjective data

The following sections analyze and discuss the results of our experiment. First, we evaluate the agreement between the subjects. We also study their bias and inconsistency during the test. Then, we analyze the impact of main factors such as the reference models (also known as source contents), the viewpoints and the animations on the obtained opinion scores and their accuracy.

4.1 Screening observers and computing mean ratings

Before starting any analysis, participants were screened using the ITU-R BT.500 recommendation [42]. Applying this procedure on our data, we did not find any outlier participants.

A common way to analyze the opinion scores of a DSIS test is to compute the Mean Opinion Score (MOS) of each stimulus.

$$MOS_e = \frac{1}{N}\sum_{i=1}^{N} s_{ie} \tag{1}$$

s_{ie} refers to the score assigned by observer i to the stimulus e. N denotes the number of subjects.

To better understand the influence of user and source variability on the opinion scores, we used the recovery model based on maximum likelihood estimation (MLE) recently introduced by Li et al. [43]. This approach recovers subjective quality scores from noisy raw measurements, by jointly estimating the subjective quality of impaired stimuli (true score), the bias and inconsistency of test subjects, and the ambiguity of the visual content all together.

$$X_{e,s} = x_e + B_{e,s} + A_{e,s} \tag{2}$$

$$B_{e,s} \sim N(b_s, v_s^2) \tag{3}$$

$$A_{e,s} \sim N(0, a_c^2) \tag{4}$$

$X_{e,s}$ is the raw opinion score. x_e is the (true) quality score of the stimulus e. $B_{e,s}$ is the noise factor of subject s on rating stimulus e, it follows a Gaussian distribution in which the mean b_s represents the subject's bias, and the variance v_s^2 represents the subject's inconsistency. The factor $A_{e,s}$ refers to the source c that corresponds to the stimulus e. Its parameter a_c^2 represents the ambiguity related to c. The estimate of each parameter (x_e, b_s, v_s, a_c) is associated with a 95% confidence interval (calculated as described in [43] [44]).

The MLE model improves classical MOS calculation by removing the uncertainty from subjects and contents. In our case, these recovered MOSs (x_e in eq. 2) remain close to classical MOSs (from eq. 1) (0.998 Spearman correlation). However, the bias, inconsistency, and content ambiguity values obtained constitute valuable information for further analysis (see paragraphs below). In the rest of this paper, we consider the recovered MOSs (x_e instead of MOS_e) as the ground truth values for our database.

4.2 Observer agreement

Before analyzing the results of the experiment, it is essential to evaluate the agreements between the subjects and whether they maintained their attentiveness during the test. To do so, we consider two types of indicators: (1) the correlations between subjects' ratings, and (2) the bias and inconsistency from the MLE model.

First, as in [21], we computed the Pearson Linear Correlation Coefficient (PLCC) and the Spearman Rank Order Correlation Coefficient (SROCC) between the scores of each observer and the MOSs of the stimuli rated by this observer. We then averaged the correlations over all the subjects. The (mean, standard deviation) of PLCC and SROCC are (0.85,0.055) and (0.81,0.063) respectively. The mean of the 2 correlations is high, while the standard deviation is rather low, which indicates a good agreement between the subjects. We then further explored the internal consistency of the subject data as proposed by [45]. For each stimulus, we randomly divided the subjects who rated it into two equal size groups (12 observers per group) and calculated the SROCC between the recovered MOSs of the 2 groups. After repeating this test 500 times, the range of correlations was found to be between 0.915 and 0.944 with a mean and a median value of 0.929. Hence, there is a high degree of inter-subject agreement despite the immersive viewing environment.

Moving to the second type of indicators, the subject's bias and inconsistency results, computed by the MLE model, are shown in Figures 4 and 5.

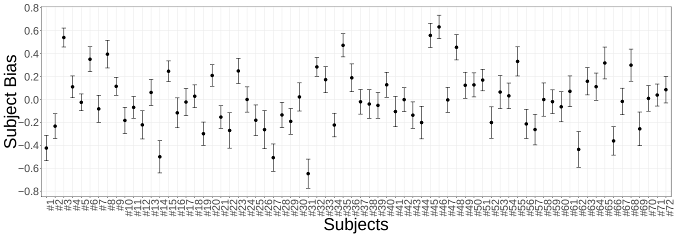


Fig. 4: Bias b_s of each subject involved in our subjective experiment, and its confidence interval.

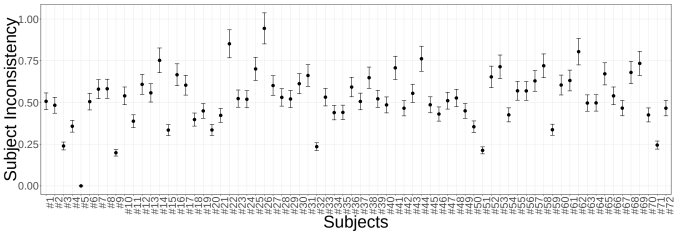


Fig. 5: Inconsistency v_s of each subject involved in our subjective experiment, and its confidence interval.

Bias reflects the sensitivity of the subject to impairments. It

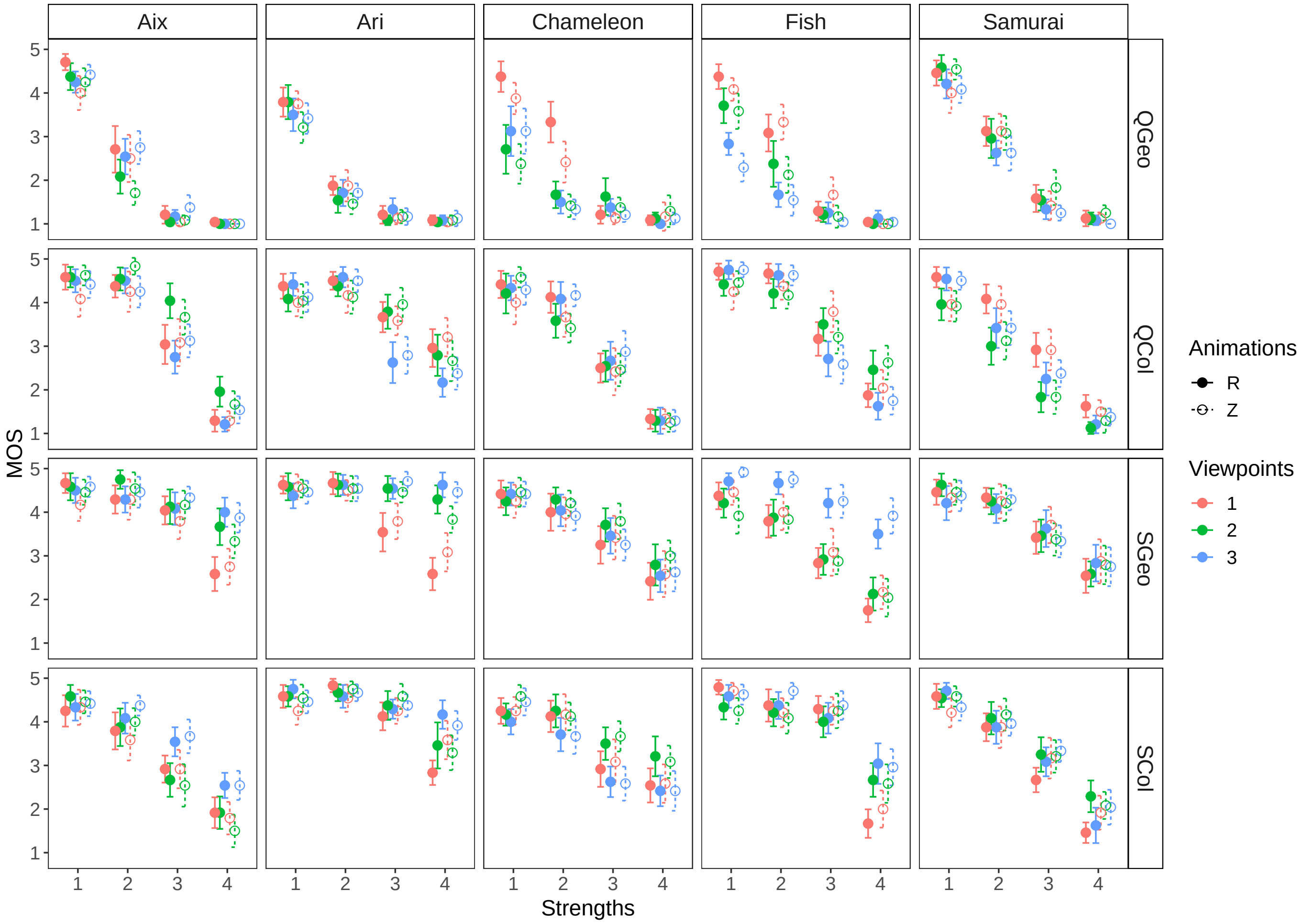


Fig. 6: The mean opinion scores of all the stimuli, associated with their confidence intervals. For a given distortion strength, the dots are horizontally spaced apart to avoid overlapping.

is a systematic error generated by the subject throughout the experiment (i.e. picky/expert participants tend to be biased toward lower scores). Inconsistency, also known as random error, points out the inattentive subjects that give random scores or subjects showing absent-mindedness for a portion of the test.

Figures 4 and 5 show that the range of bias and inconsistency values is within those of image/video experiments [43] [44] [46]. These figures reported no implausible bias or inconsistency values, nor any loose confidence intervals, which means that subjects maintained attentiveness throughout the test and were sensitive to impairments. This is coherent with the results obtained using the BT.500's outlier detection method.

Finally, we assess whether previous VR experience influences the subjects' judgments. Thus, we divided the observers into 2 groups: those who are familiar with VR (45 subjects) and those who have never tried a VR headset (27 subjects). For each group, we computed the correlations between subjects' ratings and MOS. We then averaged the correlations over all the subjects. Furthermore, we assessed the inconsistency of the 2 groups. Table 3 shows the results. We include, in the supplementary material, the corresponding boxplot of subjects's inconsistency.

TABLE 3: agreement and inconsistency of subjects familiar with VR and those without VR experience.

(Mean, SD)	PLCC	SROCC	Inconsistency
Subjects unfamiliar with VR	(0.845, 0.057)	(0.811, 0.065)	(0.536, 0.156)
Subjects familiar with VR	(0.845, 0.056)	(0.813, 0.062)	(0.517, 0.165)

For the three computed indicators, no significant difference was found in the behavior of observers with no-VR experience and those familiar with VR. We believe this is due to the fact that the task given to the participants is rather simple: observe and then vote using the trigger of the HTC Vive controller. As can be seen, no VR expertise is required, since there is no manipulation of the objects. Results also point out that our training stage was well-designed.

4.3 Factors that influence subjective opinions

Our objective is to provide a deep and evidence-based understanding of the factors that influencing subjective opinions. We quantitatively evaluate the effects of source models, distortions, viewpoints and movement on the mean opinion scores and their confidence intervals (CIs). Note that, the classic MOSs and CIs are used in this analysis since the quality scores and their corresponding CIs obtained by the MLE model are recovered from the influence of the source models (content ambiguity).

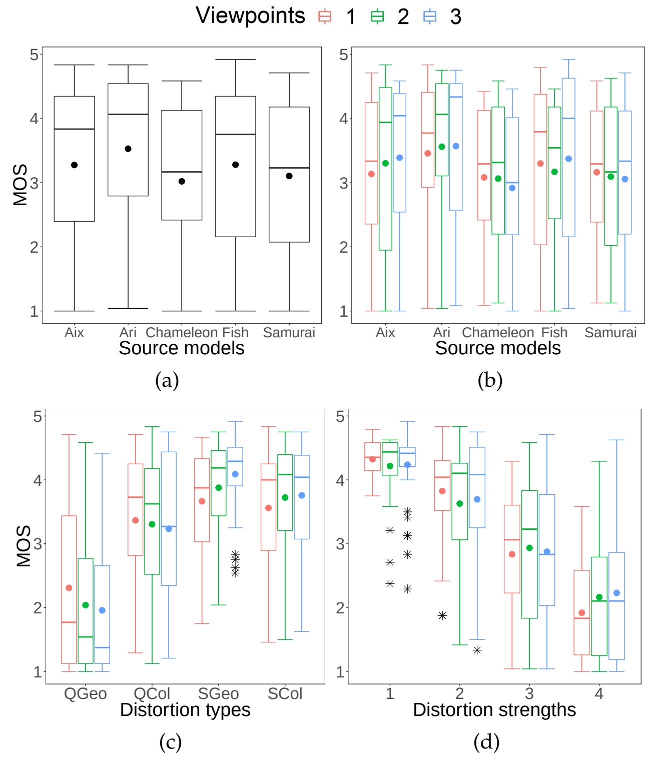
4.3.1 *Resulting MOSs and confidence intervals*

Figure 6 shows the MOSs and confidence intervals for all the stimuli, averaged over all the observers. As expected, on the whole, MOSs decrease as distortion strengths increase. We notice that observers' behavior was virtually the same for the stimuli whether they were rotating or zooming in/out. However, we can observe that the effect of the viewpoints is strongly related to the source model (e.g. Fish_SGeo vs. Chameleon_SGeo) and the distortion type (e.g. Fish_SGeo vs. Fish_SCol).

We also notice variations in confidence interval length depending on the source content (i.e. the Chameleon's CIs are globally larger than those of Ari). In addition, it seems that, overall, viewpoint *3* provides smaller CIs than viewpoint *1*. All these factors and phenomena are quantitatively analyzed in the following paragraphs. To ensure better readability in interpreting the influence of the viewpoints, we separated the results of the 2 animations. The corresponding figures are provided in the supplementary material.

4.3.2 Influence on MOSs

We ran a Multivariate Analysis of Variance (ANOVA: Reference models × distortion types × distortion strengths × viewpoints × animation types) on the rating scores of the observers. Figure 7 summarizes the most important results using boxplots of MOSs.



(a) (b)

(c) (d)

Fig. 7: Boxplots of MOSs obtained for the different factors or combination of factors. Mean values are displayed as circles.

Source models, distortion types and strengths: as expected, the ANOVA test shows that these 3 factors are the most significant factor variables (the corresponding p-values << 0.0001) (see Figure 7.a).

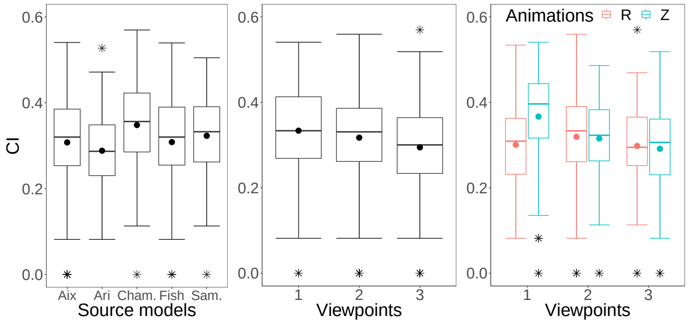
Viewpoints: there are no significant differences in the subjective scores associated with the 3 selected viewpoints (p-value=0.189). However, a significant interaction effect was found between the viewpoint and the source content (p-value << 0.0001) (see Figure 7.b). This effect appears in Figure 6. In addition, the viewpoint is also strongly related to the distortion types (p-value << 0.0001). For instance, as illustrated in Figure 7.c, viewpoint *1* is much more sensitive to geometric simplification (SGeo) than viewpoint *3*. This effect is reversed for geometric quantization (QGeo), since viewpoint *3* got the lowest average scores. Our hypothesis is that the geometry and silhouette alterations caused by QGeo are masked by rich colors and details of viewpoint *1* (viewpoint *1* is much richer than viewpoints *2* and *3*). This is not the case for geometric simplification (SGeo), which markedly degrades colors and is thus more visible on viewpoint *1*. The geometrically simplified Fish (see Figures 2 and 6) is a good case in point: observers were able to detect SGeo distortion when the stimulus was shown in viewpoints *1* and *2*. This distortion is not so apparent/visible when the Fish was displayed in viewpoint *3* and is thus harder to detect in both rotation and zoom. For QGeo, we clearly observe the opposite effect.

Figure 7.d shows that a significant interaction exists between the viewpoint and the distortion strength (with a p-value << 0.0001). Indeed, stimuli with high strength of impairment (strength=4) obtained better scores when displayed in viewpoints *2* and *3* than in viewpoint *1*. This is due to the fact that viewpoint *1* covers most of the shape and carries the most information and details. Thus, it is easier to detect loss in the visual quality of stimuli in viewpoint *1* than in viewpoints *2* and *3*. This effect is obviously less visible for high quality stimuli (strength = 1). Figure 6 shows a concrete example: for Ari geometrically simplified and shown in viewpoints *2* and *3*, as distortion forces increase, MOS values remain almost stable. These 2 viewpoints show the side and the back of the statue, respectively (see Figure 2). These areas are almost flat and contain very few geometric details/features, especially relating to the shape of the back (viewpoint *3*). Therefore, simplifying these regions, even with high strength, will not introduce introduce any markedly visible distortions to the model. This is not the case of Ari displayed in viewpoint *1*, since viewpoint *1* contains more salient features/details such as the face.

Animations: according to the ANOVA test, the animation itself does not affect significantly the perceived quality (p-value = 0.165). However, a significant interaction was found between the animation and the distortion strength (p-value < 0.0001). Indeed, weak distortions (strength = 1) are easier to detect in zoom than in rotation, while stimuli with high distortion (strength=4) obtained roughly the same score in both animations. Moreover, there is a slight interaction effect between the animation and the viewpoint (p-value=0.09). The interaction between these 2 factors will be discussed in the following subsection since it has much more influence on the CI than on the MOS.

4.3.3 Influence on confidence intervals

This time, we ran the ANOVA test on the 95% confidence intervals of the MOSs. This allows us to evaluate the impact of the factors on the dispersion of individual ratings. As above, Figure 8 summarizes the most important results using boxplots of CIs.



(a) (b) (c)

Fig. 8: Boxplots of CIs obtained for the different factors or combination of factors. Mean values are displayed as circles. In (a), *Cham.* and *Sam.* refer to Chameleon and Samurai, respectively.

Source models: When looking at Figure 8.a, it appears obvious that the source models influence the agreement among the subjects (p-value=0.0016). Indeed, selection of source models is no trivial task: some contents tend to be more difficult to rate than others. This phenomenon is represented in the MLE model by the ambiguity of the content a_c (see the supplementary material). Overall, the chameleon tends to be the source associated with the highest content ambiguity (subjects disagree). A reasonable explanation for this is that the chameleon model carries more information content than all the other models: it has a high average curvature, sharp edges, diversity of colors, and many small details to reflect its skin tone and geometrical characteristics.

Viewpoints: It is interesting to observe that the viewpoint has a significant impact on CIs (p-values=0.0035), unlike that on MOSs. Figure 8.b shows that the CIs of viewpoint *1* are larger than those of the other viewpoints. The fact that viewpoint *1* contains more details/information on color and geometry than the others implies that this viewpoint results in higher dispersion between the observers' scores.

Animations: Overall, the CIs of the stimuli in rotation are smaller than those in zoom. Still this difference is moderate (p-value=0.052). The impact of this factor is emphasized when considering the interaction between animations and viewpoints (p-value=0.0019). Indeed, Figure 8.c shows that models with animated zoom tend to be more ambiguous (result in larger CIs) than those that rotate, notably when the models are displayed in viewpoint *1*, which is the viewpoint that covers most of the shape and carries the most information. This effect can be observed in Figure 6 for Samurai, Aix and Chameleon models shown in viewoint *1* and animated with a slow zoom. This can be explained by the fact that while zooming, especially in viewpoint *1*, the observer can see more details and low-level features, which makes the task of evaluating differences between the reference and the impairment stimulus more difficult than the other HRTs.

We analyzed the ambiguity of our source contents (a_c), obtained by the MLE model, for each viewpoint and animation. The result is provided in the supplementary material. It is consistent with the findings of this section: sources with high confidence intervals are also associated with high ambiguity values.

Thus, the results point out a relationship between the ambiguity of the source (i.e. dispersion of subject ratings) and its geometric and color complexity. Models with more details are the most difficult to rate (larger confidence intervals). Furthermore, the rating is affected by the selected viewpoint: the most informative viewpoint tends to produce the largest confidence intervals, especially when combined with a zoom movement. For given distortions, an impact of the viewpoint on the MOSs can also be observed. Complex masking effects occur when considering the interaction between viewpoint and distortion. The animation has no significant impact on the perceived quality/degradation. We stress that, further studies should be carried out so that these results can be generalized to a non-VR scenario.

Our findings suggest recommendations for the design of an objective quality assessment metric for 3D meshes. First, since the models and their distortions and strengths have crucial importance on perceived quality, the metric must be able to adapt to the models, and to their shapes and colors. Moreover, it must be able to detect/capture different distortions applied on both geometry and color map. We develop such a metric in section 5. Considering the animation as a non-influential factor, it is ineligible for integration in the metric. However, since the viewpoint has an impact, albeit moderate, it may be useful to take it into account in the objective model. We investigate this in section 7.

5 Toward an Objective Metric for Assessment of Colored Mesh Quality

As outlined in the introduction, constructing an objective metric for the quality assessment of 3D content with appearance attributes is no trivial task. The main reasons are: (1) the multimodal nature of the data (geometry and color or texture information) and (2) the complex processing pipeline that constructs the final rendered image from the 3D content (computation of light-material interactions, viewpoint selection, and rasterization). To overcome this problem, we consider a data-driven approach based on the results and data of our subjective study. Thus, we propose an objective metric for colored mesh quality assessment as a linear combination of accurate geometry and color quality measurements.

5.1 Overview of our approach

The metric we propose is a full-reference multiscale metric based on curvature and color statistics computed on local corresponding neighborhoods from the original and distorted models. The metric is largely inspired by the MSDM2 frameworks from which we take the curvature features and the neighborhood correspondence mechanisms [1]. To address the color-related aspects of our metric, we consider the features introduced in the 2D image-difference framework of Lissner et al. [30]. Their color features have recently been used successfully for the quality assessment of colored 3D point clouds [31]. We refer to our metric as CMDM (for Color Mesh Distortion Measure).

Our framework is as follows: For given distorted M_{dist} and reference M_{ref} meshes, we first establish a correspondence between M_{dist} and M_{ref} (see section 5.2). Then for each scale h_i, we define a spherical neighborhood around each vertex v of M_{dist} (see section 5.3) and compute a set of local geometry and color based features over the points belonging to the neighborhood of v and their corresponding points on M_{ref} (see section 5.4). Local single-scale feature values are pooled into global multiscale features f_j. Finally, CMDM is defined as a linear combination of an optimal subset of features determined through logistic regression (see section 5.5).

5.2 Correspondence between meshes

The first objective is to establish a correspondence between the meshes being compared (M_{dist} and M_{ref}). Thus, we match each vertex v of the distorted mesh M_{dist} with its nearest 3D point $\hat{v}$ on the surface of the reference mesh M_{ref} using a fast asymmetric projection (as in MSDM2, we consider the AABB tree structure from CGAL [47]). Then, for each projected 3D point ($\hat{v}$), we compute its curvature and color using barycentric interpolation from vertices of the triangle it belongs to. This way, each vertex from M_{dist} has a corresponding point on M_{ref} (with a curvature and a color value).
The correspondence is scale-independent: it takes place once only at the beginning of the process. Nevertheless, the curvature and color values of $\hat{v}$ are updated for each scale h_i.

5.3 Neighborhood Computation

As stated above, the features used in our metric are not computed globally on the entire mesh but locally at multiple scales over spherical neighborhoods around each vertex. Thus as in [1], we define, for each scale h, a neighborhood $N(v, h)$ of radius h around each vertex v of M_{dist} as the connected set of vertices belonging to the sphere with center v and radius h. We also add to this neighborhood the intersections between this sphere and the edges of M_{dist}. The curvature and color values of the intersection points are interpolated. Then, we consider for the set of points belonging to $N(v, h)$ their projected 3D points on M_{ref} (corresponding neighborhood of $\hat{v}$). Features are computed by considering curvature and color statistics over $N(v, h) \in M_{dist}$ and $N(\hat{v}, h) \in M_{ref}$.
In this paper, we consider the following three scales: $h_i \in \{0.003BB, 0.0045BB, 0.006BB\}$, where BB is the maximum length of the Axis-Aligned Bounding Box (AABB) of the stimulus. The choice of these scales is detailed and justified in the supplementary material.

5.4 Perceptually relevant features

For each scale h, the following 8 features are computed over the local corresponding neighborhood of each vertex v of M_{dist}.

A. Geometry-based features

These features are based on mean curvature information defined at multiple scales. To compute curvature, we adopted the method developed by Alliez et al. [48], which evaluates the curvature tensor on a geodesic neighborhood around each vertex. This method is interesting and robust because it avoids the problem of sensitivity to connectivity (M_{dist} and M_{ref} do not necessarily share the same connectivity nor the same level of details). Note that, we used a radius $r = \frac{h}{3}$ for the computation of curvature as a good compromise between small radii which capture tiny details and larger radii which provide strong smoothing effects.
As in [1], we consider the following geometry features:

$$Curvature\ comparison \quad f_1^h(v) = \frac{\left\|\overline{C_v^h} - \overline{C_{\hat{v}}^h}\right\|}{max(\overline{C_v^h}, \overline{C_{\hat{v}}^h}) + k} \tag{5}$$

$$Curvature\ contrast \quad f_2^h(v) = \frac{\left\|\sigma_{C_v^h} - \sigma_{C_{\hat{v}}^h}\right\|}{max(\sigma_{C_v^h}, \sigma_{C_{\hat{v}}^h}) + k} \tag{6}$$

$$Curvature\ structure \quad f_3^h(v) = \frac{\left\|\sigma_{C_v^h}\sigma_{C_{\hat{v}}^h} - \sigma_{C_v^h C_{\hat{v}}^h}\right\|}{\sigma_{C_v^h}\sigma_{C_{\hat{v}}^h} + k} \tag{7}$$

where k is a constant to avoid instability when denominators are close to zero ($k = 1$ as in [1]). $\overline{C_v^h}$ and $\overline{C_{\hat{v}}^h}$ are Gaussian-weighted averages of curvature over the points belonging to the neighborhood $N(v, h)$ and $N(\hat{v}, h)$, respectively. Similarly, $\sigma_{C_v^h}$, $\sigma_{C_{\hat{v}}^h}$ and $\sigma_{C_v^h C_{\hat{v}}^h}$ are Gaussian-weighted standard deviations and covariance of curvature over these neighborhoods.

B. Color-based features

To compute the color features, we first transform the RGB color values of each vertex of the meshes being compared (M_{dist} and M_{ref}) into the perceptually uniform color space LAB200HL [49]. Lissener et al. recommended working in this color space since there is little cross contamination between the color attributes (lightness, chroma, hue). Each vertex v has of a lightness and two chromatic values (L_v, a_v, b_v). The chroma of the vertex is as follows: $Ch_v = \sqrt{a_v^2 + b_v^2}$.

We transposed for 3D meshes, the 2D image features proposed by [30]. These features take into account not only the luminance but also the chroma and hue components to better assess the chromatic distortions.

$$Lightness\ comparison \quad f_4^h(v) = \frac{1}{c_1(\overline{L_v^h} - \overline{L_{\hat{v}}^h})^2 + 1} \tag{8}$$

$$Lightness\ contrast \quad f_5^h(v) = \frac{\sigma_{L_v^h}\sigma_{L_{\hat{v}}^h} + c_2}{\sigma^2_{L_v^h} + \sigma^2_{L_{\hat{v}}^h} + c_2} \tag{9}$$

$$Lightness\ structure \quad f_6^h(v) = \frac{\sigma_{L_v^h L_{\hat{v}}^h} + c_3}{\sigma_{L_v^h}\sigma_{L_{\hat{v}}^h} + c_3} \tag{10}$$

$$Chroma\ comparison \quad f_7^h(v) = \frac{1}{c_4(\overline{Ch_v^h} - \overline{Ch_{\hat{v}}^h})^2 + 1} \tag{11}$$

$$Hue\ comparison \quad f_8^h(v) = \frac{1}{c_5\overline{\Delta H_{v\hat{v}}^h}^2 + 1} \tag{12}$$

where $\overline{L_v^h}$, $\overline{L_{\hat{v}}^h}$, $\overline{Ch_v^h}$ and $\overline{Ch_{\hat{v}}^h}$ denote the Gaussian-weighted averages of Lightness and Chroma computed respectively over the set of points belonging to $N(v,h)$ and $N(\hat{v},h)$. $\sigma_{L_v^h}$, $\sigma_{L_{\hat{v}}^h}$ and $\sigma_{L_v^h L_{\hat{v}}^h}$ are Gaussian-weighted standard deviations and covariance of lightness in the mentioned neighborhood. The term $\overline{\Delta H_{v\hat{v}}^h}$ refers to the Gaussian-weighted average hue difference between $N(v,h)$ and $N(\hat{v},h)$. It is defined as follows: $\Delta H_{v\hat{v}} = \sqrt{(a_v - a_{\hat{v}})^2 + (b_v - b_{\hat{v}})^2 - (Ch_v - Ch_{\hat{v}})^2}$. The constants c_1, c_2, c_3, c_4 and c_5 were set respectively to 0.002, 0.1, 0.1, 0.002 and 0,008 as in [30].

We invert the scaling of the color-based features so that they are consistent with curvature-based features (i.e. each color feature $f_j^h = 1 - f_j^h$). This way, a value of 0 means that there is no local (geometric and color) distortion around vertex v. All features ∈ [0, 1].

5.5 Global perceptual quality score

The set of local geometric and color features, presented in the subsection above, is computed for each vertex of the distorted mesh and for each scale h_i. The local multiscale measure of the features is simply the average of its single-scale values.

$$f_j(v) = \frac{1}{n}\sum_{i=1}^{n} f_j^{h_i}(v) \tag{13}$$

where n is the number of scales used. It is defined in section 5.3 as well as h_i the scale values used (neighborhood radii) .

We aim to obtain a global score of visual distortion according to each feature (f_j). So, we average the local values of each feature over all the vertices.

$$f_j = \frac{1}{|M_{dist}|}\sum_{v \in M_{dist}} f_j(v) \tag{14}$$

where $|M_{dist}|$ is the number of vertices of the distorted mesh. The features f_j are all within the range [0, 1].

Our metric is then defined as a combination of the features f_j. However, choosing the best combination model is a crucial problem. For prediction of the color-image differences [30], the authors used a factorial combination model, while Meynet et al. considered a linear model for their point cloud quality metric [31]. In our case, we chose to consider a linear model: (1) to make the optimization easier and (2) because we tried nonlinear models such as Minkowski pooling, which did not provide better performance. Thus, the global multiscale distortion (GMD) score is computed as follows:

$$GMD_{M_{dist} \to M_{ref}} = \sum_{j \in S} w_j f_j \tag{15}$$

S is the set of feature indexes of our linear model. w_j weights the contribution of each feature to the overall distortion prediction. $GMD_{M_{dist} \to M_{ref}}$ evaluates the distortion of the distorted model regarding the reference model. In order to strengthen the robustness of our method and to obtain a symmetric measure, we also compute $GMD_{M_{ref} \to M_{dist}}$ and we retain the average as the final distortion measure $CMDM$.

$$CMDM = \frac{GMD_{M_{dist} \to M_{ref}} + GMD_{M_{ref} \to M_{dist}}}{2} \tag{16}$$

As in [50], the optimal subset of features of $CMDM$ and their corresponding weights are obtained through an optimization computed by logistic regression. The optimization is based on cross-validation, using the ground truth dataset from our subjective experiment (see section 6.3).

6 Results and evaluation

In this section, we evaluate the performance of our metric and compare it to state-of-the-art approaches, including 2D image metrics. To train and evaluate our metric, we used the ground truth database obtained from our subjective experiment (section 3). In this section, we do not take into account either the influence of the viewpoints or that of the animations: for a given stimulus, we averaged its recovered MOSs over the 3 viewpoints and the 2 animations. Thus, the database used is composed of 80 stimuli. We also validate our metric on a dataset from [5], composed of distorted textured meshes.

6.1 Performance evaluation measures

In order to evaluate the performance of objective metrics, we compare the predicted quality scores given by these metrics to the ground truth subjective data. The standard performance evaluation measure consists in computing the Pearson Linear Correlation Coefficient (PLCC) and the Spearman Rank Order Correlation Coefficient (SROCC) between the metric predictions and subjective scores (MOS). These indices measure, respectively, the accuracy and the monotonicity of the predictions. Note that, the Pearson correlation (PLCC) is computed after a logistic regression which provides a non-linear mapping between the objective and subjective scores. This allows the evaluation to take into account the saturation effects associated with human senses.

However, the correlations ignore the uncertainty of the subjective scores. Therefore, as a complementary assessment of the performance of the objective metrics, we also implement the framework recently proposed by Krasula et al. [51]. This methodology consists in determining the classification abilities of the metrics according to two scenarios:

- (A) Different vs. Similar: this analysis assesses how well can the metric distinguish between significantly different and similar pairs of stimuli. The first step consists in determining the pairs in the dataset rated significantly different. To do so, we conduct a statistical test (t-test) on the raw subjective scores. Then for each pair of stimuli (i, j), we compute the absolute difference of the predicted scores ($|\Delta_{\text{PredictedScores}}(i,j)|$) and measure how well these values are able to correctly classify the pairs of stimuli.
- (B) Better vs. Worse: this analysis is performed on the significantly different pairs only. The significantly different pairs (i, j) are divided into 2 groups: i better than j (($\Delta_{\text{MOS}}(i,j) > 0$)) and i worse than j (($\Delta_{\text{MOS}}(i,j) < 0$)). We then measure, according to ($\Delta_{\text{PredictedScores}}(i,j)$) values, how well the metric is able to predict this classification.

As can be seen, these scenarios take into account the uncertainty of the subjective scores. Both scenarios refer to a binary classification problem (different/similar and better/worse). As a performance indicator, we consider the Receiver Operating Characteristic (ROC) and, more precisely, the Area Under the Curve (AUC) values. AUC is a direct indicator of the performance/ability of the classifiers (1.0 corresponds to a perfect classification, 0.5 corresponds to a random one). In what follows, it is noted by AUC_{DS} and AUC_{BW} for scenarios A and B, respectively.

6.2 Single feature prediction performance

This section evaluates the prediction performance of each feature implemented in our multiscale metric. Table 4 shows the correlations of the individual features with the recovered MOSs, as well as their classification abilities.

TABLE 4: Performance of individual features.

Feature	Id	*PLCC*	*SROCC*	AUC_{DS}	AUC_{BW}
Curvature comparison	f_1	0.5	0.44	0.6	0.75
Curvature contrast	f_2	0.45	0.43	0.59	0.73
Curvature structure	f_3	0.3	0.32	0.53	0.67
Lightness comparison	f_4	0.58	0.69	0.69	0.83
Lightness contrast	f_5	0.7	0.71	0.7	0.87
Lightness structure	f_6	0.68	0.71	0.69	0.87
Chroma comparison	f_7	0.38	0.59	0.64	0.78
Hue comparison	f_8	0.33	0.43	0.6	0.71

Overall, the best features are those based on the lightness information (especially f_5, f_6). They correlate well with the subjective scores and provide a good performance in identifying the significantly different stimuli as well as the stimuli of better quality. For the geometry-based features, f_1 and f_2 perform better than f_3. However, this does not necessarily point to the ineffectiveness of f_3. Finally, regarding the chromatic feature, chroma comparison f_7 seems more relevant than hue comparison f_8. Note that, the geometry features are penalized by the color quantization (QCol), since this type of distortion is applied only on the vertex colors and does not affect the model geometry at all. Removal of this distortion improves their performance, notably with respect to correlations (increase to 0.7 for f_1 anf f_2). This latter analysis is available in the supplementary material.

6.3 Toward an Optimal Combination of features

Our metric contains 8 different features f_j. In this 8 dimensional space, some features are obviously more significant than others. Also, features may be redundant with one another, and if all the features are taken into account, this could potentially lead to an overfitting. Therefore, in the same vein as [50], we conduct two Leave-One Out Cross-Validation tests (LOOCV) on the data obtained from our subjective experiment to select an optimal subset of features. Each cross-validation test divides the database into a training set that serves to optimize feature weights using linear regression and a test used for testing the obtained metric.

1) We split the training and test sets according to the source models. Given that there are 5 sources in our database, we leave 1 source model and its distortions out for testing, while the remaining stimuli (4 models * 16 distorted stimuli) are used for training. Thus after 5 folds, each source model has been used as a test set.
2) Similar to test 1, but we divide the database according to the distortion types (regardless of the model). We train the metric on 3 distortion types out of 4 (5 models * 12 distorted stimuli) and test on the fourth type. After 4 folds, each distortion type has been used once for testing.

These 2 types of LOOCV tests provide a good measure of the robustness of our metric. We exhaustively search through all possible combinations of features (255 combinations), and select the feature-subset that generates the best average performance of $CMDM$ over all the test sets (9 folds) in terms of the mean of PLCC and SROCC. We obtained that the final model of our metric is composed of only 4 features: Curvature contrast (f_2), Lightness contrast (f_5) and structure (f_6) and chroma comparison (f_7). The optimal features found are consistent with the results of the single feature performance. The results of our metric and comparisons with state-of-the-art approaches are reported in the following sections.

6.4 Comparisons of objective metrics

In this section, we present the results of the cross-validation tests, described in the previous subsection. As an ablation study, we compare our metric with two of its versions trained with different subsets of features: *CMDM_Geo* that takes into account only the geometry features and *CMDM_Col* based only on color features. As a baseline, we also include results of a classical color distance *D_LAB*, which is the average of the color difference (in LAB2000HL) computed symmetrically between the reference and the distorted model. Finally, we compare our metric with 3 state-of-the-art full-reference image quality metrics (IQMs): *SSIM* [34], *HDR-VDP2* [52], *iCID* [53]. To apply these IQMs, we generate for each 3D object in our database, a set of 18 snapshots taken from different viewpoints (camera positions regularly sampled). The global quality score of a stimulus, given by an IQM, is then the average of the objective scores over all its snapshots. The parameters of IQMs, as well as snapshots of the camera positions are provided in the supplementary material.

Figure 9 compares the overall performance of the tested metrics for the 2 cross-validation scenarios presented in 6.3. Tables 5 and 6 detail the results of each test set.

For the LOOCV test according to the source models, Figure 9 demonstrates that *CMDM* outperforms other model-based metrics. It shows almost the same performance as IQMs in terms of correlations and detection of better quality stimuli (AUC_{BW}). IQMs provide better results in identifying the significantly different pairs of stimuli (AUC_{DS}). We believe this is primarily related to the advantage of IQMs over our metric and other model-based metrics regarding their natural incorporation/knowledge of the entire rendering pipeline. Indeed, IQMs operate on snapshots that consider the same rendering, apparent brightness and lighting conditions as those seen by participants. On the contrary, our metric only considers 3D data, without any knowledge of the rendering conditions. Considering the LOOCV test among the distortions, we notice that our metric performs better than the others, including IQMs. The color-based version of our metric (*CMDM_Col*) also produces good results. IQMs show a significant decrease in performance, compared to the LOOCV based on source models. These observations corroborate previous results by Lavoué et al. [35]: image-based metrics perform very well when evaluating the quality of different versions of a single source, yet they are less accurate when differentiating/ranking distortions applied

TABLE 5: Performance comparison of several metrics in a cross-validation test among source models

	Aix				**Ari**				**Chameleon**				**Fish**				**Samurai**			
	PLCC	*SROCC*	AUC_{DS}	AUC_{BW}	*PLCC*	*SROCC*	AUC_{DS}	AUC_{BW}	*PLCC*	*SROCC*	AUC_{DS}	AUC_{BW}	*PLCC*	*SROCC*	AUC_{DS}	AUC_{BW}	*PLCC*	*SROCC*	AUC_{DS}	AUC_{BW}
CMDM	**0.958**	**0.956**	0.783	0.982	0.96	0.91	0.823	0.986	0.83	0.83	0.692	0.943	0.93	0.914	0.805	0.987	0.933	0.944	0.746	0.976
CMDM_Geo	0.53	0.621	0.562	0.79	0.68	0.468	0.577	0.788	0.457	0.474	0.504	0.76	0.554	0.554	0.622	0.788	0.462	0.407	0.598	0.737
CMDM_Col	0.778	0.791	0.793	0.913	0.491	0.553	0.633	0.799	0.764	0.788	0.631	0.914	0.941	0.903	0.866	0.99	0.76	0.779	0.631	0.887
D_LAB	0.791	0.826	0.77	0.924	0.282	0.497	0.523	0.737	0.776	0.747	0.609	0.897	0.734	0.779	0.713	0.9	0.546	0.659	0.59	0.787
SSIM	0.896	0.909	0.782	0.959	0.973	0.932	**0.924**	0.993	0.823	0.868	0.683	0.951	**0.959**	0.929	0.9	**0.992**	0.957	0.915	0.846	0.989
HDR-VDP2	0.893	0.853	0.728	0.958	**0.976**	**0.947**	0.877	**0.998**	0.849	0.818	0.727	0.963	0.895	0.897	0.751	0.981	**0.978**	0.962	0.86	0.995
iCID	**0.958**	0.932	**0.849**	**0.983**	0.953	0.929	0.85	0.989	**0.924**	**0.921**	**0.743**	**0.986**	0.954	**0.935**	**0.912**	0.988	0.966	**0.968**	**0.914**	**0.999**

TABLE 6: Performance comparison of several metrics in a cross-validation test among distortion types

	QGeo				**QCol**				**SGeo**				**SCol**			
	PLCC	*SROCC*	AUC_{DS}	AUC_{BW}	*PLCC*	*SROCC*	AUC_{DS}	AUC_{BW}	*PLCC*	*SROCC*	AUC_{DS}	AUC_{BW}	*PLCC*	*SROCC*	AUC_{DS}	AUC_{BW}
CMDM	0.882	0.825	0.537	0.933	**0.917**	**0.924**	**0.893**	**0.973**	**0.93**	**0.94**	**0.871**	**0.995**	0.841	**0.841**	0.641	**0.939**
CMDM_Geo	0.686	0.481	0.597	0.8	0.121	0.288	0.457	0.373	0.596	0.73	0.638	0.874	0.455	0.486	0.486	0.745
CMDM_Col	0.787	0.758	0.493	0.904	0.821	0.845	0.772	0.943	0.889	0.908	0.838	0.984	**0.853**	0.795	**0.712**	0.934
D_LAB	0.653	0.677	0.501	0.826	0.799	0.851	0.72	0.932	0.765	0.841	0.702	0.938	0.598	0.629	0.613	0.832
SSIM	0.875	0.794	0.709	0.903	0.792	0.708	0.696	0.896	0.722	0.756	0.623	0.901	0.588	0.704	0.598	0.855
HDR-VDP2	**0.946**	**0.938**	**0.805**	**0.987**	0.882	0.78	0.724	0.939	0.736	0.762	0.59	0.9	0.767	0.777	0.632	0.918
iCID	0.88	0.838	0.632	0.926	0.86	0.81	0.785	0.939	0.738	0.761	0.645	0.906	0.631	0.747	0.657	0.872

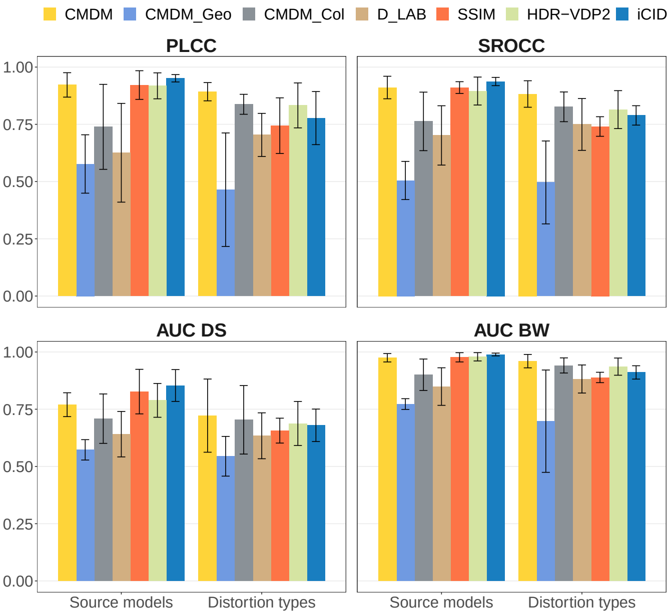


Fig. 9: Performance comparison of several metrics on two cross-validation tests. Mean performance evaluation measures are reported. Error bars indicate the standard deviation over the test sets.

on different sources. More details are provided in Tables 5 and 6.

From Table 5, it can be seen that our metrics and IQMs demonstrate a good stability over the models. The performance (especially the correlations) of the *D_LAB* and *CMDM_Col* metrics drop dramatically for the Ari model. We also notice that the quality of the Chameleon model was hardest to predict, since almost all the metrics (except *D_LAB* and *CMDM_Col*) exhibit a poorer performance than the other models. This is coherent with our findings in section 4.3.3.

When considering each distortion type separately (Table 6), several observations can be made. First, our metric performs very well on 3 types of distortion out of 4: For QCol, SCol and SGeo, it outperforms significantly the other metrics, and particularly IQMs. However, our metric shows a poor performance when distinguish between similar and different pairs corrupted by geometric quantization (QGeo). For this distorsion, *HDR-VDP2* performs significantly better in terms of correlations and classification abilities. *CMDM* seems to underestimate the impact of geometric quantization (QGeo), which is particularly harmful for such high-resolution models in our database. We believe that this is due to the fact that this distortion superimposes the vertices of the stimulus, meaning that we cannot know or control exactly which vertex color is taken into account in Unity's import and render pipelines. This case points out an advantage for image-based quality metrics and highlights the importance of taking rendering into account in the assessment of visual quality.

6.5 Recommended weights

To provide the recommended model of our metric, we averaged the weights obtained for each training subset of the two LOOCV tests. CMDM is thus defined, for the three selected scales ($h_i \in \{0.003BB, 0.0045BB, 0.006BB\}$), as follows:

$$CMDM_{rec} = 0.091f_2 + 0.22f_5 + 0.032f_6 + 0.656f_7 \quad (17)$$

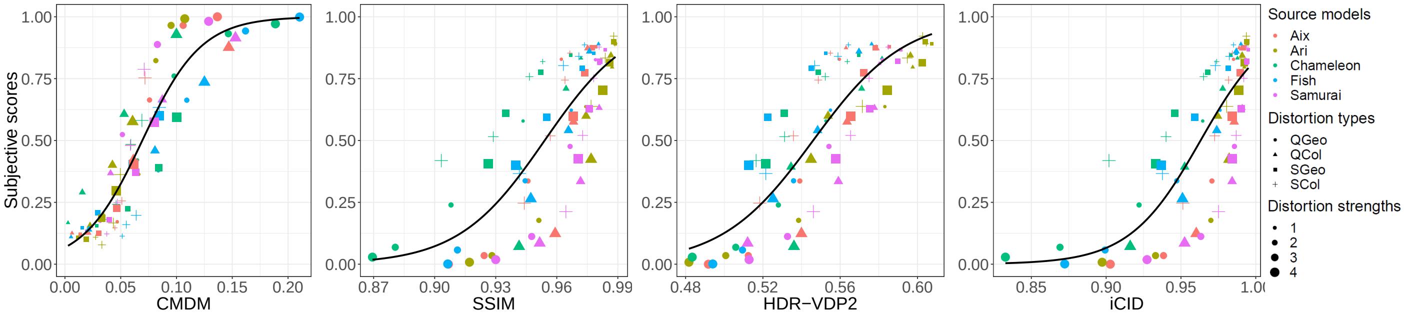
In order to reveal the relative importance of each of the 4 features, we scaled the weights presented in the equation above with the standard deviation of the features. Scaled weights are 0.333, 0.46, 0.07 and 0.136, respectively, for f_2, f_5, f_6 and f_7. The curvature and lightness contrast features (f_2 and f_5) have the highest overall importance. It would seem that users are particularly sensitive to artifacts that harm the contrast (both geometric and color contrasts).

We evaluate the performance of the tested metrics, including $CMDM_{rec}$, on the whole dataset (80 stimuli). The results are reported in Table 7. Figure 10 shows the subjective scores with respect to objective metric values.

TABLE 7: Performance comparison of different metrics on the whole dataset.

	PLCC	*SROCC*	AUC_{DS}	AUC_{BW}
CMDM$_{rec}$	**0.913**	**0.9**	**0.782**	**0.968**
CMDM_Geo	0.501	0.437	0.604	0.749
CMDM_Col	0.745	0.746	0.732	0.893
D_LAB	0.55	0.603	0.651	0.805
SSIM	0.797	0.799	0.716	0.912
HDR-VDP2	0.853	0.84	0.703	0.944
iCID	0.825	0.83	0.747	0.924

CMDM performs notably better than the others in terms of correlations. Moreover, the AUC values reflect its good



(a) (b) (c) (d)

Fig. 10: Scatter plot of subjective scores versus objective metric values for all the dataset. Each point represents one stimulus. The fitted logistic function is displayed in black.

classification abilities in both Different vs. Similar and Better vs. Worse analyses. This shows the good robustness of our metric: it is able to differentiate and rank stimuli from different sources and different distortions.

6.6 Validation on a dataset of textured 3D meshes

To evaluate the robustness of our recommended metric (eq. 17) and to verify that it did not just learn the distortions that are specific to our dataset, we tested $CMDM_{rec}$ on a new dataset. Only few subject-rated datasets of 3D models with attributes are available to the scientific community [5] [22]. We consider the LIRIS Textured Mesh Database [5], produced from a subjective study based on a pairwise comparison method. This database is composed of 136 textured meshes, obtained from 5 source models subjected to texture and geometry distortions. Indeed, the authors generated 20 distorted versions of each source. They also selected a model (the Dwarf) among the 5, and associated it with 36 mixed distortions (combination of geometry and texture distortions). As each of these 6 subsets of the database was rated separately, they cannot be assessed together. To evaluate the robustness of our approach, we selected the most difficult subset, namely the one containing mixed distortions. The source model of this subset is a scan of a *Dwarf* statue that has been scanned and reconstructed into a textured mesh of 250004 vertices. Distortions are combinations of 3 geometry distortions (geometric quantization, simplification, smoothing), each applied with 2 strengths and 2 texture distortions (JPEG compression, sub-sampling), each applied with 3 strengths. As can be seen, these compound distortions differ significantly from the distortions generated in our dataset. Before applying our metric, we transferred the texture color information into vertex colors (we generate the vertex color by picking the corresponding color from the texture).

The results are summarized in Table 8. We include results of the IQMs presented previously, as well as the results obtained by Guo et al. [5] for different metrics: three metrics applied on rendered videos of the stimuli (the Discrete Cosinus Transform-based (*DCT*) metric [54], the *PSNR* and the *MS-SSIM* [55] applied on all frames and averaged) and three metrics directly applied on textured meshes (*FQM* [4] based on a weighted combination of two simple mesh and texture image error measures, CM_1 and CM_2 [5] both defined as a linear combination of mesh quality and texture quality). Note that, Table 8 shows only the correlation measures since subjective scores are derived from a paired-comparison method and are therefore not associated with CIs. In the supplementary material, we illustrate the subjective scores with respect to the values of the tested metrics.

TABLE 8: Performance comparison of different metrics on a new dataset. For metrics marked with a *, the values are reprinted from [5].

	PLCC	SROCC
CMDM_{rec}	**0.862**	**0.872**
SSIM	0.624	0.657
HDR-VDP2	0.83	0.844
iCID	0.502	0.552
Video-DCT*	0.32	0.50
Video-PSNR*	0.33	0.58
Video-MS-SSIM*	0.67	0.66
FQM*	0.64	0.66
CM_1*	0.74	0.77
CM_2*	0.80	0.85

Our metric provides the best results, although it was trained on a different dataset presenting different sources and different distortions and even a different color representation. *SSIM* and *iCID* show poor performances. They may be affected by the fact that the snapshots used do not have the same rendering and lighting conditions as those of the experiment. Note that results for *SSIM*, computed using snapshots of the stimuli, are consistent with those reported by [5], which are computed on the rendered videos used in the subjective test.

Our metric also outperforms CM_2, which represents the state-of-the-art of textured mesh quality assessment, and which was learned on similar data. This metric is a global combination of mesh and texture distortion measures (*MSDM2* and *MS-SSIM*, respectively). This tends to validate the fact that operating fully on the mesh domain (like our metric) ensures a better performance than combining errors computed on different domains (i.e., mesh and texture image). These results also confirm the great robustness of our metric compared to IQMs.

7 Integration of the Viewpoint

According to the findings of our subjective experiment, the viewpoint of stimuli may have a significant impact on user quality assessment. Thus, we hypothesized that incorporating this factor into our objective metric should improve its results. Indeed, the invisible parts of the 3D model do not contribute to its visual appearance. Given a stimulus and a camera position, we determined, in a preprocessing step, which vertices are visible and which vertices are occluded by other faces of the mesh (using ray-vertex intersections). Note that, we have ignored the slight change in visibility of vertices on the borders/boundaries of objects caused by their animation. Thus, our objective metric is now computed only over the visible vertices. We can redefine Equation 14 as follows:

$$f_i = \frac{1}{|M'_{dist}|} \sum_{v \in M_{dist}} f_i(v)\Psi(v) \qquad (18)$$

where Ψ is a function that returns 0 or 1 according to the visibility of vertex v and $|M'_{dist}|$ is the number of visible vertices of the distorted mesh.

To evaluate the performance of the new metric $CMDM_{vis}$, we used a subset of our database consisting of 240 stimuli. Indeed, for a given stimulus, we considered its 3 viewpoints and averaged the recovered MOSs of the 2 animations. We tested the performance of the metric with and without integrating the visibility on these 240 stimuli. Similarly for the IQMs, we considered 2 scenarios: (1) without taking the visibility into account, so we computed the IQMs on multiple snapshots taken from different viewpoints and (2) computing the IQMs directly on the snapshot taken from the real viewpoint displayed to the observer (IQM_{vis}). Table 9 shows the improvement/evolution of metric results when incorporating the viewpoint. The improvement is defined as the difference in the evaluation measures (correlations and AUC) computed before and after integrating the viewpoint (e.g. for a given metric M: $\Delta PLCC = PLCC_{M_{vis}} - PLCC_M$). The full results of the tested metrics in both scenarios are provided in the supplementary material.

TABLE 9: Performance evolution of different metrics before and after integrating the viewpoint

	$\boldsymbol{\Delta PLCC}$	$\boldsymbol{\Delta SROCC}$	$\boldsymbol{\Delta AUC_{DS}}$	$\boldsymbol{\Delta AUC_{BW}}$
CMDM	0	-0.005	-0.001	0
SSIM	0.018	0.03	0.025	0.012
HDR-VDP2	-0.022	0.018	-0.053	0.001
iCID	0.057	0.071	0.049	0.03

We obtained, through the 2 versions of all the metrics, roughly the same performance in terms of correlations and classification abilities (no significant performance improvement). Our hypothesis is that this lack of improvement is due to the fact that only a small subset of the dataset is actually rated significantly different for its different viewpoints. This led us to conduct a more precise study: we identified those stimuli with viewpoints associated with significantly different subjective scores. We found out that the viewpoint has a significant influence only on 88 pairs out of 22695 pairs of stimuli rated significantly different.

Thus, instead of considering all the possible pairs of stimuli (240*239/2), we compared each stimulus separately according to its 3 viewpoints VP (e.g. *Fish_SGeo_4_VP1* vs. *Fish_SGeo_4_VP2*, *Fish_SGeo_4_VP1* vs. *Fish_SGeo_4_VP3* and *Fish_SGeo_4_VP2* vs. *Fish_SGeo_4_VP3*). This limited the study to 240 pairs of stimuli (80 stimuli*3 possible combinations of pairs of viewpoints), 88 of which were significantly impacted by the viewpoints. The results are shown in Table 10. Note that, only the AUC values are reported since this study is based on pairs of stimuli and thus the correlations could not be computed.

TABLE 10: Performance comparison of different metrics on the pairs of stimuli significantly affected by the viewpoints.

	CMDM$_{\text{vis}}$	**SSIM$_{\text{vis}}$**	**HDR-VDP2$_{\text{vis}}$**	**iCID$_{\text{vis}}$**
AUC_{DS}	**0.602**	0.56	0.561	0.58
AUC_{BW}	0.58	0.66	**0.8**	0.668

Without integrating the viewpoint information, the AUC values of all the metrics are equal to 0.5. Including the viewpoint slightly improved the results. Still, this improvement is low, except for *HDR-VDP2$_{vis}$*, which showed a good ability to recognize the stimulus of higher quality in the pair. This study takes the first step toward integrating the knowledge of the viewpoint into objective metrics. The fact that the IQMs exhibited a relatively poor performance, even though they were computed directly on the displayed view, shows that it is considerably difficult to distinguish the perceived quality of different viewpoints of the same 3D model. Further work is still needed to produce efficient metrics in this difficult scenario. In particular, we hypothesize that classical pooling should be replaced by more sophisticated pooling. It could also be useful to consider visual attention models.

8 Conclusion and Future work

In this work, we designed and produced a large subjectively-rated database of colored 3D meshes. This database is composed of 480 dynamic stimuli and obtained through a subjective study based on the DSIS method, in a virtual reality environment. The stimuli were generated from 5 source models subjected to geometry and color distortions. Each stimulus was associated with 6 HRTs: combinations of 3 viewpoints and 2 animations. This study allowed us to draw interesting conclusions regarding the masking effects that occur when considering the interaction between viewpoint and distortion. Although animation, by itself, has a moderate impact on subjects' opinions and CIs, the impact of this factor is emphasized when it interacts with other factors such as distortion strength for subjective scores and viewpoint for CIs. Moreover, results show that the ambiguity of the source is potentially related to its geometric and color complexity. The more visible the content's information/complexity (as in zoom for example), the higher the ambiguity.

We developed a perceptually-validated full-reference metric *CMDM* for evaluating the quality of colored 3D meshes. To achieve this, we adapted a set of perceptually-relevant curvature-based and color-based features. We further show how to select an optimal subset of features and use them to train the metric (LOOCV tests using a ground truth dataset). Extensive evaluation shows that *CMDM* provides good results and good stability in terms of correlations and classification abilities. It also demonstrates

a good robustness: *CMDM* is able to differentiate and rank stimuli from different sources and different distortions, unlike IQMs which perform very well when assessing the quality of different versions of a single source, but are less accurate when ranking distortions applied on different sources. Last but not least, we demonstrate that our metric can also be used for textured meshes.
Our ground truth database, subjective scores, and the metric code are made publicly available online.

As future work, we plan to further explore how to effectively incorporate visibility information into objective measures. We would also like to produce a huge subject-rated database of 3D models, in order to be able to envisage the creation of end-to-end deep-learning approaches. Finally, we will work towards adapting the perceived quality of the objects according to the position/movement of the subject in the VR scene.

Acknowledgments

This work was supported by French National Research Agency as part of ANR-PISCo project (ANR-17-CE33-0005).

References

[1] G. Lavoué, "A Multiscale Metric for 3D Mesh Visual Quality Assessment," *Computer Graphics Forum*, vol. 30, no. 5, pp. 1427–1437, 2011.

[2] L. Váša and J. Rus, "Dihedral Angle Mesh Error: a fast perception correlated distortion measure for fixed connectivity triangle meshes," *Computer Graphics Forum*, vol. 31, no. 5, 2012.

[3] K. Wang, F. Torkhani, and A. Montanvert, "A Fast Roughness-Based Approach to the Assessment of 3D Mesh Visual Quality," *Computers & Graphics*, 2012.

[4] D. Tian and G. AlRegib, "Batex3: Bit allocation for progressive transmission of textured 3-d models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 23–35, 2008.

[5] J. Guo, V. Vidal, I. Cheng, A. Basu, A. Baskurt, and G. Lavoue, "Subjective and objective visual quality assessment of textured 3D meshes," *ACM Transactions on Applied Perception*, vol. 14, no. 2, 2016.

[6] N. Ye, H.-P. Seidel, K. Myszkowski, R. Mantiuk, A. Steed, K. Wolski, P. Didyk, R. K. Mantiuk, and D. Giunchi, "Dataset and Metrics for Predicting Local Visible Differences," *ACM Transactions on Graphics*, vol. 37, no. 5, pp. 1–14, 2018.

[7] G. Lavoué and R. Mantiuk, "Quality assessment in computer graphics," *Visual Signal Quality Assessment: Quality of Experience (QoE)*, pp. 243–286, 2015.

[8] B. Watson, "Measuring and predicting visual fidelity," *ACM Siggraph*, pp. 213–220, 2001.

[9] B. E. Rogowitz and H. Rushmeier, "Are image quality metrics adequate to evaluate the quality of geometric objects?" *Proceedings of SPIE*, pp. 340–348, 2001.

[10] Y. Pan, I. Cheng, and A. Basu, "Quality metric for approximating subjective evaluation of 3-D objects," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp. 269–279, apr 2005.

[11] G. Lavoue, E. Drelie Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi, "Perceptually driven 3D distance metrics with application to watermarking," in *SPIE*, vol. 6312, 2006.

[12] G. Lavoué, "A local roughness measure for 3D meshes and its application to visual masking," *ACM Transactions on Applied Perception*, vol. 5, no. 4, 2009.

[13] M. Corsini, E. D. Gelasca, T. Ebrahimi, and M. Barni, "Watermarked 3-D Mesh Quality Assessment," *IEEE Transactions on Multimedia*, vol. 9, no. 2, pp. 247–256, feb 2007.

[14] S. Silva, B. S. Santos, C. Ferreira, and J. Madeira, "A Perceptual Data Repository for Polygonal Meshes," *2009 Second International Conference in Visualisation*, pp. 207–212, jul 2009.

[15] K. Christaki, E. Christakis, and P. Drakoulis, "Subjective Visual Quality Assessment of Immersive 3D Media Compressed by Open-Source Static 3D Mesh Codecs," in *25th International Conference on MultiMedia Modeling (MMM)*, 2018, pp. 1–12.

[16] Y. Nehmé, J. P. Farrugia, F. Dupont, P. Le Callet, and G. Lavoué, "Comparison of subjective methods, with and without explicit reference, for quality assessment of 3D graphics," in *ACM Conference on Applied Perception*, 2019.

[17] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," in *SPIE Optical Engineering + Applications*, 2018, p. 18.

[18] Evangelos Alexiou, I. Viola, T. M. Borges, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A comprehensive study of the rate-distortion performance in MPEG point cloud compression," *APSIPA Transactions on Signal and Information Processing*, vol. 8, 2019.

[19] L. A. d. S. Cruz, E. Dumic, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation : Towards a definition for test conditions," *International Conference on Quality of Multimedia Experience*, 2019.

[20] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Point Cloud Rendering after Coding : Impacts on Subjective and Objective Quality," *arXiv:1912.09137*, pp. 1–13, 2019.

[21] H. Su, Z. Duanmu, W. Liu, Q. Liu, and Z. Wang, "Perceptual Qualty Assessment of point Clouds," *IEEE International Conference on Image Processing*, pp. 3182–3186, 2019.

[22] E. Zerman, C. Ozcinar, P. Gao, and A. Smolic, "Textured Mesh vs Coloured Point Cloud : A Subjective Study for Volumetric Video Compression," *International Conference on Quality of Multimedia Experience*, 2020.

[23] K. Vanhoey, B. Sauvage, P. Kraemer, and G. Lavoué, "Visual quality assessment of 3D models: On the influence of light-material interaction," *ACM Transactions on Applied Perception*, vol. 15, no. 1, 2017.

[24] F. Torkhani, K. Wang, and J.-m. Chassery, "A Curvature Tensor Distance for Mesh Visual Quality Assessment," 2012.

[25] M. Corsini, M. C. Larabi, G. Lavoué, O. Petrik, L. Váša, and K. Wang, "Perceptual Metrics for Static and Dynamic Triangle Meshes," *Computer Graphics Forum*, vol. 32, no. 1, pp. 101–125, feb 2013.

[26] I. Abouelaziz, A. Chetouani, M. El Hassouni, L. J. Latecki, and H. Cherifi, "No-reference mesh visual quality assessment via ensemble of convolutional neural networks and compact multi-linear pooling," *Pattern Recognition*, vol. 100, p. 107174, 2020.

[27] Z. C. Yildiz, A. C. Oztireli, and T. Capin, "A machine learning framework for full-reference 3D shape quality assessment," *Visual Computer*, vol. 36, no. 1, pp. 127–139, 2020.

[28] G. Nader, K. Wang, H. Franck, and F. Dupont, "Just Noticeable Distortion Profile for Flat-Shaded 3D Mesh Surfaces," *IEEE Trans. on Visualization and Computer Graphics*, 2016.

[29] J. Guo, V. Vidal, A. Baskurt, and G. Lavou, "Evaluating the local visibility of geometric artifacts," *ACM Symposium in Applied Perception*, 2015.

[30] I. Lissner, J. Preiss, P. Urban, M. S. Lichtenauer, and P. Zolliker, "Image-difference prediction: From grayscale to color," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 435–446, 2013.

[31] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "PCQM: A Full-Reference Quality Metric for Colored 3D Point Clouds," in *International Conference on Quality of Multimedia Experience*, 2020.

[32] S. Yang, C.-H. Lee, and C. Kuo, "Optimized mesh and texture multiplexing for progressive textured model transmission," *ACM Multimedia*, pp. 676–683, 2004.

[33] F. Caillaud, V. Vidal, F. Dupont, and G. Lavoué, "Progressive compression of arbitrary textured meshes," *Computer Graphics Forum*, vol. 35, no. 7, 2016.

[34] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[35] G. Lavoue, M. C. Larabi, and L. Vasa, "On the Efficiency of Image Metrics for Evaluating the Visual Quality of 3D Models," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 8, pp. 1987–1999, 2016.

[36] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," pp. 209–216, 1997.

[37] H. Lee, G. Lavoué, and F. Dupont, "Rate-distortion optimization for progressive compression of 3D mesh with color attributes," *The Visual Computer*, vol. 28, no. 2, pp. 137–153, may 2012.

[38] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, pp. 1427–41, 06 2010.

[39] S. Ling, J. Gutiérrez, K. Gu, and P. Le Callet, "Prediction of the influence of navigation scan-path on perceived quality of free-viewpoint videos," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 204–216, 2019.

[40] ITU-T P.910, " Subjective video quality assessment methods for multimedia applications," *International Telecommunication Union*, 2009.

[41] G. Regal, R. Schatz, J. Schrammel, and S. Suette, "VRate: A Unity3D Asset for integrating Subjective Assessment Questionnaires in Virtual Environments," pp. 1–3, 05 2018.

[42] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures BT Series Broadcasting service," *International Telecommunication Union*, 2012.

[43] Z. Li and C. Bampis, "Recover subjective quality scores from noisy measurements," pp. 52–61, 04 2017.

[44] Z. Li, C. Bampis, L. Janowski, and I. Katsavounidis, "A simple model for subject behavior in subjective experiments," *arXiv:2004.02067v2*, 2020.

[45] M. Chen, Y. Jin, T. Goodall, X. Yu, and A. C. Bovik, "Study of 3d virtual reality picture quality," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 89–102, 2020.

[46] P. Pérez, N. García, and . Villegas, "Subjective assessment of adaptive media playout for video streaming," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019, pp. 1–6.

[47] P. Alliez, S. Tayeb, and C. Wormser, "Aabb tree.cgal 3.5 edition," 2009.

[48] P. Alliez, D. Cohen-Steiner, O. Devillers, B. Lévy, and M. Desbrun, "Anisotropic polygonal remeshing," *ACM Trans. Graph.*, vol. 22, no. 3, p. 485–493, Jul. 2003.

[49] I. Lissner and P. Urban, "Toward a unified color space for perception-based image processing," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1153–1168, 2012.

[50] Y. Liu, J. Wang, S. Cho, A. Finkelstein, and S. Rusinkiewicz, "A no-reference metric for evaluating the quality of motion deblurring." *ACM Transactions on Graphics*, vol. 32, no. 6, pp. 171–175, 2013.

[51] L. Krasula, K. Fliegel, P. Le Callet, and M. Klíma, "On the accuracy of objective image and video quality models: New methodology for performance evaluation," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, pp. 1–6.

[52] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, Jul. 2011.

[53] J. Preiss, F. Fernandes, and P. Urban, "Color-image quality assessment: From prediction to optimization," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1366–1378, 2014.

[54] F. Xiao *et al.*, "Dct-based video quality evaluation," *Final Project for EE392J*, vol. 769, 2000.

[55] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, 2003, pp. 1398–1402 Vol.2.

Yana Nehmé received a double degree in electronics and digital technology from the faculty of engineering of the Lebanese University and Ecole polytechnique de l'Université de Nantes. She is now a PhD student at the Institut National des Sciences Appliquées de Lyon (Insa Lyon) under the supervision of Prof. Guillaume Lavoué. Her research interests include visual perception and visual quality assessment, immersive data, augmented and virtual reality.

Florent Dupont is full professor at the Université Claude Bernard Lyon 1 in France. His main research concerns 3D digital image processing and computational geometry within the ORIGAMI Computer Graphics team of LIRIS, a joint research unit to CNRS (UMR 5205), INSA Lyon, Université Claude Bernard Lyon 1, Université Lumière Lyon 2 and Ecole Centrale de Lyon.

Jean-Philippe Farrugia is an associate professor in Université Lyon 1 and LIRIS CNRS laboratory. He is co-leader and co-founder of the ORIGAMI computer graphics team in this lab. His research interests are realistic and real time rendering, visual perception, augmented and virtual reality.

Patrick Le Callet received both a M.Sc. and a PhD degree in image processing from Ecole polytechnique de l'Université de Nantes, where he is now a full professor. He passed the "Aggrégation" (credentialing exam) in electronics of the French National Education at the Ecole Normale Superieure de Cachan. He is one of the seven members of the steering board of the LS2N CNRS lab and scientific director of the cluster "Ouest Industries Créatives". He is mostly engaged in research dealing with the application of human vision modeling in image and video processing, cognitive computing, Quality of Experience assessment, visual attention modeling and applications, perceptual video coding and immersive media processing. He is co-author of more than 250 publications and communications and co-inventor of 16 international patents on these topics. He serves or has been served as associate editor or guest editor for several Journals such as IEEE TIP, IEEE STSP, IEEE TCSVT, SPRINGER EURASIP Journal on Image and Video Processing, SPIE JEI, IEEE IVMSP-TC and IEEE MMSP-TC and one the founding members of EURASIP SAT (Special Areas Team).

Guillaume Lavoué (M'11-SM'13) is a full professor at the École Nationale d'Ingénieurs de Saint-Étienne (ENISE), affiliated to University of Lyon. His current research interests lie in the areas of Computer Graphics and Virtual Reality, including geometry processing, 3D graphics compression, applied visual perception, and objective/subjective quality assessment. Guillaume Lavoué authored more than 30 journal papers and has been chairing the IEEE SMC TC on Human Perception and Multimedia Computing (2013-2017).