CHAPTER 18

Subjective and objective quality assessment for volumetric video

Evangelos Alexiou[a], **Yana Nehmé**[b], **Emin Zerman**[c,g], **Irene Viola**[a], **Guillaume Lavoué**[b], **Ali Ak**[d], **Aljosa Smolic**[f], **Patrick Le Callet**[d], **and Pablo Cesar**[a,e]

[a]DIS, Centrum Wiskunde en Informatica, Amsterdam, the Netherlands
[b]Origami, LIRIS, Lyon University, Lyon, France
[c]STC Research Center, Mid Sweden University, Sundsvall, Sweden
[d]IPI, LS2N, Nantes University, Nantes, France
[e]Multimedia Computing, TU Delft, Delft, the Netherlands
[f]V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

As discussed in previous chapters, immersive video technologies create visual content for human consumption, as they "attempt to emulate a real world through a digital or simulated recreation" [1]. The reconstructed volumetric video can be viewed from any angle, providing 6 degrees-of-freedom (DoF) interaction capabilities and is suitable for extended reality (XR) applications, e.g., augmented reality (AR) or virtual reality (VR) applications (see Part 5: Applications). However, the increased level of interactivity offered to the user comes at the cost of a vast amount of data that needs to be processed in a radically different way with respect to traditional video, which in turn prompted further scientific research and active involvement of the MPEG and JPEG standardization bodies [2,3]. Considering that human viewers are the end-users, a thorough understanding of the human visual system (HVS) is necessary to ensure high quality of experience (QoE), as discussed in Chapter 1.

Visual quality assessment is critical to ensure the highest QoE in media technologies. Volumetric video might undergo distortions during processing, compression, transmission, and rendering, which negatively affect the fidelity of the original content. This creates a need for mechanisms that quantify these distortions; that is, new methods to capture subjective quality scores and predict the perceived quality of the content displayed to the viewer. Such mechanisms can be helpful in selection of optimal schemes and tuning of parameters in perceptual terms to improve the QoE. For instance, estimated quality scores are commonly employed to optimize the efficiency of content delivery systems by increasing the effectiveness of compression and transmission methods, considering the trade-off between quality and data size.

[g] Emin Zerman was with V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland at the time of writing this chapter.

Immersive Video Technologies
https://doi.org/10.1016/B978-0-32-391755-1.00024-9

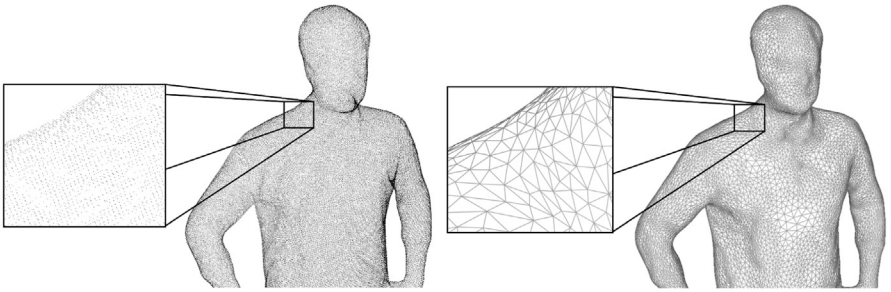

Figure 18.1 Primary elements defining the topology of two commonly used representation methods for volumetric video: point cloud and mesh. Points in 3D space for a point cloud are shown on left, and vertices plus edges for a mesh are shown on right. See Fig. 18.2 for the rendering with color information.

The problem of visual quality assessment has been well-studied for traditional video. Decades of studies on how to assess the quality brought many standards and recommendations, which detail test methodologies, experimental designs, and evaluation procedures for reproducible subjective quality experiments [4–6]. Similarly, many objective quality metrics were developed for the assessment of traditional image and video quality [7]. Despite the breadth of video quality estimation solutions for traditional video, extending the methodologies and algorithms devised for traditional media formats to immersive contents is not trivial. For instance, new subjective quality assessment methods are required to accommodate the higher DoF and to imitate real-life consumption of such richer imaging modalities. Furthermore, since the volumetric video is represented by data types, inherently different than pixels on a regular grid, corresponding objective quality metrics are designed differently.

Volumetric contents are most frequently delivered in the form of 3D polygonal meshes, or 3D point clouds [8]. For a 3D polygonal mesh, the model shape is defined by a set of vertices accompanied by connectivity information to form polygons (typically triangles), whereas the color information for the polygonal (or triangular) faces is determined through texture maps. A 3D point cloud representation is defined by a set of points placed on 3D space, with point coordinates determining the shape, and associated color attributes the color of the model, as shown in Fig. 18.1. Hereafter, we refer to them simply as meshes and point clouds. Both of them have unique characteristics and pose different challenges for visual quality assessment.

Volumetric video is essentially a collection of 3D models, which are played back at a certain frame rate, which gives the viewer the illusion of a continuous movement. This is the same principle still used in traditional video. As the advances and findings in image quality assessment studies were useful for traditional video quality, the findings for static 3D models will bring insight into the quality assessment of dynamic sequences that form

volumetric video. Therefore, in this chapter, we cover subjective and objective quality assessment methods for both static and dynamic 3D models, represented as both point clouds and meshes. In particular, we provide the following:

- An overview of subjective quality assessment methodologies with respect to the mode of inspection (i.e., non-interactive and interactive)
- A descriptive list of publicly available subjectively annotated datasets
- A summary of user studies that compare different parameters in the design of subjective quality experiments
- An overview of objective quality metrics grouped per operating principle (i.e., model-based and image-based)
- A unified table with publicly available objective quality metrics
- Advantages and disadvantages of different objective quality assessment approaches

18.1. Subjective quality assessment

As the quality is defined as a subjective trait in the context of multimedia signal processing [9], the golden standard to obtain quality values for the volumetric video is to conduct subjective user studies. Nevertheless, subjective experiments are resource and time expensive to conduct, as they require careful experiment design, expertise on the subject matter, a dedicated space to conduct the user study, and participants' and experimenters' time. Although the objective quality metrics can provide estimated quality scores much faster, subjective evaluations are crucial in quality assessment, as they provide ground truth data for further research and development.

Commonly, subjective experiments are designed according to standards recommended by standardization communities [4,5] or expert groups formed by researchers [6]. The standardization efforts for immersive imaging technologies are still underway. Recently, a new methodology has been standardized for the evaluation of omnidirectional content [10,11]; however, this only takes into account rotational movements in 3DoF, which are not suitable for volumetric videos. Work items currently under study in International Telecommunication Union (ITU) involve subjective methodologies for interactive VR [12] and QoE assessment of XR tele-meetings [13]. Nevertheless, currently, there are no specific standards or recommendations for the new volumetric video.

The existing recommendations and standards for traditional images and video do not take 6DoF interaction into account, as viewers are essentially passive spectators. In other words, in traditional imaging, the viewers can see the entire visual stimulus (i.e., images or video) whenever they are looking at the display. Once they are presented the stimulus, they can inspect it for a fixed duration (e.g., around 10 seconds is a common choice [4]). After this duration, they are asked to either vote on the quality or select the preferred stimulus, depending on the adopted test methodology.

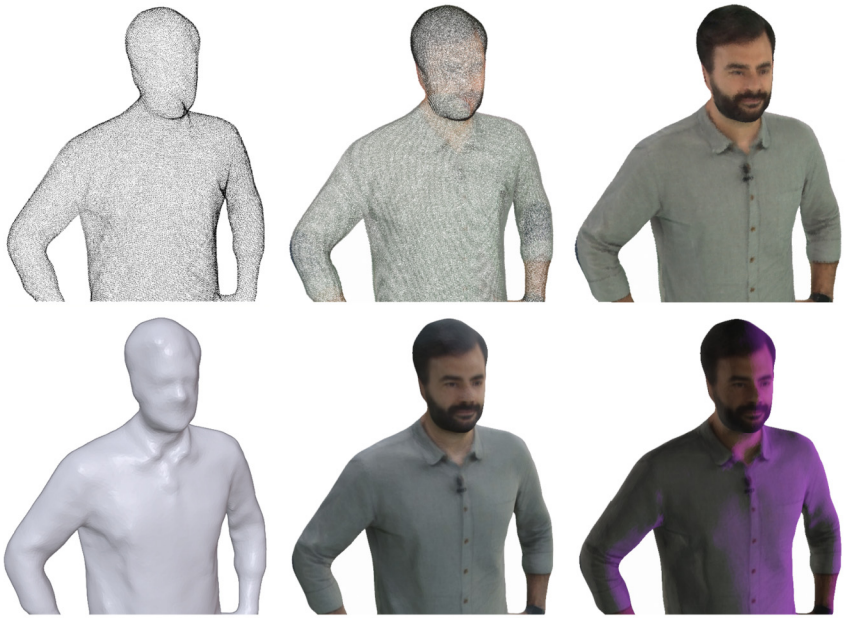

Figure 18.2 Illustration of different aspects affecting the perception of volumetric content: colorless vs colored, point size for point clouds, and lighting for meshes. Top row refers to point cloud representation: (top-left) colorless point cloud, (top-center) colored point cloud with point size 1, (top-right) colored point cloud with point size 3. Bottom row refers to mesh representation: (bottom-left) colorless surface, (bottom-center) textured surface, (bottom-right) textured surface rendered with lighting.

For volumetric video, seeing the whole content at once is not possible, as the content is occluding itself at any given time due to its 3D nature. To ensure that the collected subjective quality scores are representative for the whole volumetric video, the experimenter needs to ensure that the volumetric video is inspected properly by the subjective experiment participants. There are two main ways to facilitate this: let viewers interact with the volumetric video themselves or present a representative stimulus (e.g., a sequence of images from predefined viewpoints) in which the viewer's interaction is non-existent. The former methods better simulate real-life use cases of 3D media consumption, whereas the latter approaches provide the same experience across subjects granting reproducibility, and enable utilization of well-established practices that have been developed for evaluation of traditional video.

There are many aspects that need to be taken into account while designing and conducting subjective user studies, in addition to the user interaction aspect discussed above. The volumetric video can be represented with different 3D formats (e.g., meshes and point clouds), which may contain different attributes; these models can be colorless,

Table 18.1 Various aspects in subjective user studies for volumetric video.

Aspects	Variables
3D Representation	Mesh, Point Cloud
Temporal Variation	Static, Dynamic
Attributes	Colorless, Colored
Mode of Inspection	Passive, Interactive
Methodology	Single stimulus, Double stimulus, Multiple stimulus, Pairwise comparison
Distortion Types	Noise, Compression, Simplification, Smoothing, Sub-sampling, Transmission error
Rendering Parameters	Lighting, Background, Point size (for point clouds)
Display Devices	2D monitors, 3D monitors, Head-mounted displays, Smartphones

colored, or textured. Moreover, there are different rendering parameters that should be configured, which affect the appearance of the models. In Fig. 18.2, examples are presented regarding the effect of the lighting on textures meshes and the point size on point clouds, which, if not carefully chosen, may lead to the appearance of holes (small point size) or patchy areas (large point size). Subjective experiments may also use different display devices or evaluation methodologies to collect quality scores. These experiments make use of either static or dynamic 3D models, and they might feature various types of practical distortions; such as additive or multiplicative noise, mesh simplification or point cloud sub-sampling, compression, transmission, and smoothing. A summary of all these aspects can be seen in Table 18.1.

In this section, we categorize the scientific efforts on subjective quality assessment with respect to the mode of inspection, which dictates if and how the viewer will interact with the visual stimulus. In each subsection, we further divide the studies with respect to the 3D representation.

18.1.1 Non-interactive user studies

One of the ways to collect subjective quality scores for volumetric video is to limit the user interaction completely (i.e., the viewer cannot interact with the stimulus). In this case, a certain visualization technique is used to create visual stimuli, and the volumetric video is represented to the participants of the subjective test in the same manner. There might be various approaches in creating the visual stimuli, which mostly create traditional images or a traditional video that is composed of rendered images of the volumetric video content.

A common way to create these video sequences is to select a certain camera trajectory (e.g., a camera rotating around the object, while staying on the horizontal plane and looking towards the object, as shown in Fig. 18.3). Rendering images from these

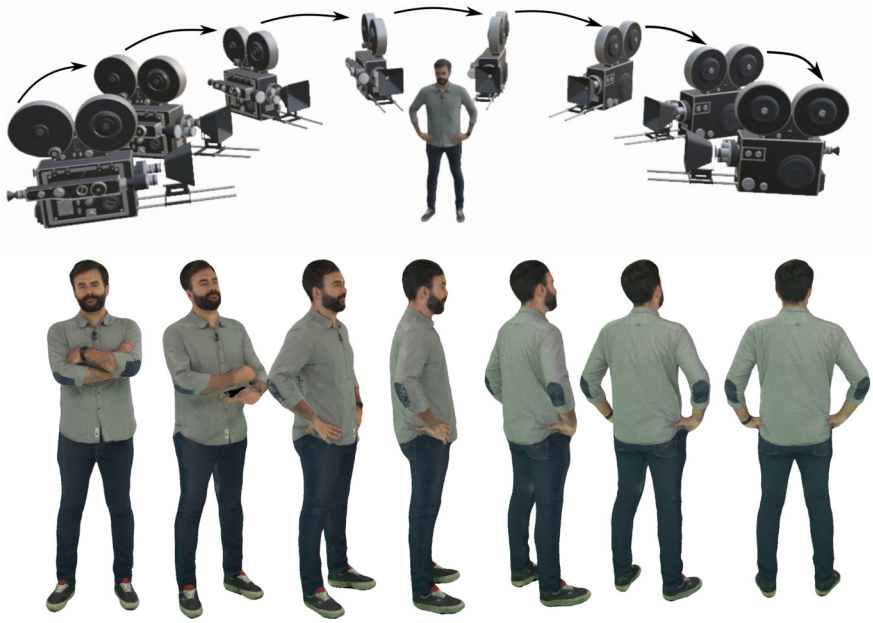

Figure 18.3 One common method to create videos for subjective experiments with no user interaction is to create traditional videos of the volumetric video sequences with a fixed camera trajectory, as shown above. Sample renderings of such a video are shown below. Attribution: The 3D camera model in the figure is created by Jesse Johnson and is licensed under Creative commons attribution (CC BY 4.0).

camera locations and sequencing them as video frames will create a traditional video, which shows the volumetric video from many different viewpoints. Following their creation, these video sequences are shown to the subjective test participants, who in turn determine the quality of the volumetric video sequence. In this setup, the viewer does not and cannot interact with the volumetric video itself, which means that the observer cannot change the viewpoint and can only watch the pre-rendered video. Therefore the user is passive in this scenario.

Although this lack of interaction is a disadvantage for the viewer, it is also an advantage for the experimenter, since this lack of interaction also removes any inter-viewer variation that might occur due to interactivity, i.e., they ensure that the same stimuli will be seen by all participants. Moreover, these approaches do not require complex rendering software or devices, since the resulting still images or videos can be easily visualized on 2D screens using commodity software, and they minimize external biases and conditions that can influence the final results, such as novelty effect or cybersickness.

In the following subsections, we discuss how the subjective user studies are conducted in the literature without user interaction, after grouping them per 3D content representation (e.g., point clouds and meshes).

18.1.1.1 User studies for point clouds

The majority of non-interactive experiments in the literature of point cloud quality assessment, focuses on static contents [14–21]. When evaluating static contents in a non-interactive manner, the experimenter does not have to worry about possible interaction effects between the camera movement and the sequence actions. For example, by simply using circular camera paths, the content will be visualized by all angles without missing information regarding the visual quality of occluded regions. Passive inspection has also been used for dynamic contents [22–24], using predefined camera paths to ensure the same user experience. Moreover, different types of distortions have been studied in the literature: from noise in geometric and textural information of point clouds [15,25–27], to compression artifacts [14–20,22], rendering approaches [19–21,27] and adaptive streaming algorithms [24]. By default, point clouds lead to the perception of holes; hence, the size of points is typically configured so as to enable visualization of watertight objects. In most studies, the experimenters assign the same point size for an entire model, whereas in some experiments different point sizes are assigned based on local densities [14,28,29]. Finally, double stimulus variants denote the most popular evaluation methodologies.

Javaheri et al. [14] evaluate static models representing inanimate objects and human figures at three quality levels, using a spiral camera path moving around a model from a full view to a closer look to capture images from different perspectives. Cubic geometric primitives of adaptive size based on local neighborhoods were employed for rendering purposes. Animated sequences were created and rated by subjects, using the sequential double stimulus impairment scale (DSIS) methodology. The color attributes of the models remained uncompressed to assess the impact of these geometry-only degradations. Su et al. [15] employed a virtual camera orbiting around the point cloud models at a fixed distance to capture the views, which were displayed using the DSIS methodology with simultaneous visualization of the reference and distorted stimuli. A wide set of colored models were distorted using different types of degradations, including Gaussian noise in both topology and texture, octree down-sampling, and compression artifacts from the MPEG test models. More recently, Lazzarotto et al. [16] conducted a crowd-sourced user study to evaluate compression distortions from both conventional and learning-based codecs. The point clouds were rendered using splats of adaptive size, with the camera orbiting at a fixed distance.

Conducting experiments in a single laboratory setting can make questions arise about the generalizability of the results. Even when standardized procedures are adopted to select and screen users, there might be human-related biases that can influence the

results. For this reason, cross-laboratory testing helps in checking the validity and in corroborating the outcomes of the test. Cruz et al. [17] report the results of a subjective evaluation campaign that was issued in the framework of the JPEG Pleno [3] activities. Subjective experiments were conducted in three laboratories for quality assessment of colored point clouds under an octree- and a projection-based encoding scheme, using the DSIS methodology. A different camera path was defined per content, and a fixed point size was specified per model and degradation level. This is reported to be the first study making use of contents consumed outer-wise (e.g., objects, human figures) and inner-wise (e.g., scenes). Perry et al. [18] report experiments that were performed in four independent laboratories that participated in the relevant JPEG exploration study activities, evaluating the performance of V-PCC and G-PCC using colored point clouds. The DSIS with side-by-side visualization was employed using fixed-size point primitives to display the models. The experimental setup of each laboratory varied; yet, the collected subjective scores exhibited high inter-laboratory correlation.

In some early studies, conversion to mesh was considered as an alternative for rendering point clouds before being assessed. The degradations under test were applied on the raw point clouds; subsequently, a surface reconstruction algorithm was applied to render the 3D models. Javaheri et al. [27] performed subjective evaluation of de-noising algorithms against impulse noise and Gaussian noise. Video sequences of the reference and the degraded models after reconstruction using the screened Poisson method [30] were sequentially shown to human subjects, before rating the visual quality of the latter. Alexiou et al. [19] present a study where subjective experiments were conducted in five test laboratories to assess the visual quality of colorless point clouds using the same reconstruction algorithm as a rendering methodology. The point clouds were degraded using octree-pruning, and the observers evaluated the mesh models under a simultaneous DSIS methodology. Although different 2D monitors were employed, subjective scores were found to be strongly correlated among the participated laboratories. In [20], the same dataset was evaluated under various 3D display types/technologies (i.e., passive, active, and auto-stereoscopic).

Different point-based and mesh rendering methodologies were employed for subjective quality evaluation by Javaheri et al. [21]. The point clouds were distorted by geometric compression artifacts and rendered using three approaches: colorless point primitives of fixed size with shading; point primitives of fixed size, rendered using the original texture information without shading; and colorless meshes obtained after screened Poisson surface reconstruction, with shading. Each rendering approach was evaluated in separate sessions, using the sequential DSIS methodology.

The aforementioned studies focused on evaluating static contents under various degradations and rendering setups. Few studies have been involved with evaluating dynamic sequences using non-interactive approaches. Schwarz et al. [2] evaluate both static and dynamic colored point cloud models under several encoding categories, settings,

and bitrates, from a quality assessment campaign that was conducted in the framework of the call for proposals issued by the MPEG committee [31]. The contents were rendered using cubes as primitive elements of fixed size; animated image sequences of the models captured from predefined viewpoints were generated and assessed using the absolute category rating (ACR) methodology. Subjective evaluations of a volumetric video data set that was acquired and released in the context of the study was performed by Zerman et al. [23], under compression artifacts from the MPEG V-PCC. The point clouds were rendered using primitive ellipsoidal elements of fixed size, and animated sequences were generated from predefined camera paths. The stimuli were subjectively assessed using two methodologies; that is, a side-by-side evaluation with DSIS and a pairwise comparison.

The previous studies focused on evaluating a single point cloud content at the time, which, depending on the context, might not be realistic. For example, several objects can be placed and viewed at the same time in a virtual space, or in a real-time communication scenario, multiple people could be present in the same scene. In [24], subjective quality assessment of dynamic, colored point clouds is conducted in an adaptive streaming scenario hosted by the system described in [32], in which more than one models is placed in the same scene under different arrangements, and were visited with different navigation paths. The streamed cues were subjectively evaluated using an ACR methodology in a desktop setting. For the purposes of the study, volumetric video sequences were selected and encoded at different quality levels using V-PCC. Among the experimental parameters, different bandwidth conditions, bitrate allocation schemes, and prediction strategies were examined.

18.1.1.2 User studies for meshes

Pioneering subjective quality tests involving meshes were conducted on still rendered images [33,34]; those two early studies both assessed the visual impact of simplification artifacts and concerned static and geometry-only meshes; the same applies in [35], which evaluated the impact of geometry compression. Subsequent passive interaction experiments considered meshes with color/texture attributes [36–41] or dynamic meshes [8,42,43]. These studies are detailed below. They considered different types of stimuli (still images or videos) and different protocols (ACR, DSIS, pairwise comparison), adapting existing image/video methodologies for passive inspection of 3D models by observers: 2D still images or generated videos of animated models.

Watson et al. [33] used still rendered images to evaluate the visual impact of mesh simplification using the DSIS methodology. Váša and Rus [35] and Doumanoglou et al. [38] also used still rendered images in their studies to evaluate the impairment of geometry compression and the visual impact of geometry and texture resolution on the quality of textured human body reconstructed meshes, respectively. Both of these studies considered a pairwise comparison methodology. Rogowitz et al. [34] conducted

two subjective quality assessment experiments: the first involved 2D static images of simplified 3D objects, whereas the second was performed on rendered videos of these objects in rotation. The results showed that lighting conditions have a strong influence on perceived quality and that observers perceive the quality of still images and animations differently. The authors concluded that the quality of 3D objects cannot be correctly assessed using static 2D projections (still images may mask the effect of light and shading), and thus it is important that the object moves.

Based on these findings, some researchers allowed users to interact freely and in real time with the model by rotating and zooming it, as detailed in Section 18.1.2.2. However, others decided to control the viewpoints visualized by the user showing an animation of the 3D object to avoid cognitive overload that can alter human judgments.

Guo et al. [37] opted for this approach to assess the influence of lighting, shape, and texture on the perception of artifacts for textured meshes. They animated each object in the dataset with a low-speed rotation and generated videos that were displayed to observers during the test. The subjective study was conducted using the pairwise comparison method. The same experimental procedure was adopted by Vanhoey et al. [44] to investigate the impact of light-material interactions on the perception of geometric distortion of 3D models.

Pan et al. [36] conducted a subjective experiment on textured meshes to assess the perceptual interactions between the geometry and color information. They considered only geometry and texture sub-sampling distortions. They animated their meshes with a slow rotation and the experiment was based on the DSIS methodology. Nehmé et al. [39] provide the first public dataset for meshes with vertex colors produced in VR. The dataset was obtained through a subjective study based on the DSIS methodology. The stimuli were rendered at a fixed viewing distance from the observer in a virtual scene, under different viewpoints and animated in real-time with either slow rotation or slow zoom-in. The study allowed to analyze the impact of several factors, such as viewpoints and animations on both quality scores and their confidence intervals.

The above works considered quality assessment of static meshes. Váša and Skala [42] and Torkhani et al. [43] were the first authors to propose quality assessment experiments involving dynamic meshes (without color/texture). Distortions included diverse types of noise and compression. The rated stimuli were videos of mesh sequences, rendered from fixed viewpoints. Used methodologies were respectively single stimulus rating [43] and multiple stimulus rating [42].

More recently, Zerman et al. [8] considered the ACR with Hidden Reference (ACR-HR) methodology to compare dynamic textured meshes and colored point clouds in the context of a VV compression scenario utilizing the appropriate state-of-the-art compression techniques for each 3D representation. They built a database and collected user quality opinion scores using rendered version of VVs, shown to the participants on an LCD display.

The previous experiments were conducted in laboratories, under controlled environments and with high-end equipment. Along with laboratory subjective experiments, crowd-sourcing experiments have become very popular in recent years, especially during COVID-19 pandemic, where participants could not be physically present in the lab. However, conducting subjective quality assessment tests in a crowd-sourcing setting imposes several challenges, notably those related to the lack of control over the participants' environment and the reliability of the participants, since the latter are not supervised. A recent study was conducted to investigate whether a crowd-sourcing test can achieve the accuracy of a laboratory test for 3D graphics [45]. For this purpose, the authors designed a crowd-sourcing experiment that replicates as much as possible the lab experiment presented in [40], which was conducted in VR. Specifically, they used the same dataset of 3D models and the same experimental methodology (i.e., DSIS). Since in crowd-sourcing the test environment cannot be fully controlled, videos of rotating stimuli were displayed to the participants to limit their interactions with the 3D objects. The results of this study showed that under controlled conditions and with a proper participant screening approach, a crowd-sourcing experiment based on the DSIS method can be as accurate as a laboratory experiment. It is worth mentioning that crowd-sourcing is quite faster to evaluate large datasets, yet the most time intensive task is building and designing the experimental framework (or setup) (user-friendly tool, control viewer environment, add screening test, etc.). Based on these findings, a large-scale crowd-sourcing experiment was conducted to rate the perceived quality of the largest dataset of textured meshes to this date [41]. This dataset allowed to analyze the impact of the distortions and model characteristics (geometric and color complexity) on the perceived quality of textured meshes.

18.1.2 Interactive user studies

An alternative way to collect subjective user quality scores is to conduct user studies in more interactive experimental settings, which account for more realistic scenarios of consumption for 3D content. Considering that there are no recommendations for interactive protocols, most of the efforts make use of well-established methodologies (e.g., ACR or DSIS) in experimental setups ranging from desktop arrangements to XR applications, which accommodate interactivity with varying DoF.

In all cases, the contents are placed in a virtual scene, designed by the experimenter depending on the task at hand (i.e., background, lighting, etc.). Moreover, the users are given the means to handle the camera position and orientation at any given moment to inspect the contents under evaluation at their will. For instance, in desktop setups the contents are displayed on flat-screen monitors with user interactions typically being registered through the mouse cursor or computer keyboards. In XR settings, the contents are visualized through a head-mounted display (HMD), with the users controlling their viewpoint either by physical movements in the real world, or by controllers.

By design, interactive evaluation protocols lead to individual visual experiences across users, which ultimately affect their opinion regarding the visual quality of the content under inspection. However, it is advocated that such methodologies are better adjusted to the interactive nature of richer imaging modalities, with user quality scores inherently containing the preferred type of interaction. To compensate this uncertainty in user ratings, it is common for the experimenters to either recruit more participants or to allow interactions without imposing time limitations. Finally, enabling interactive protocols allows the experimenter to analyze the behavior of users with 3D visual data, and explore inter-dependencies between interactivity patterns and perception of quality.

In the following subsections, we discuss subjective user studies using interactive evaluation protocols reported in the literature, clustered per 3D content representation.

18.1.2.1 User studies for point clouds

The interactive user studies for point clouds generally use three different platforms: desktop devices [25,28,29,46–49], AR [26], and VR [50–52] headsets. Many studies focus on static point clouds [25,26,28,29,46–48,50], whereas only a couple of the studies focus on dynamic point clouds [49,51]. Similarly to passive inspection experiments, studies on both colorless [25,26,46] and colored [28,29,47–52] models have been conducted, while different types of distortions and point size selection strategies have been employed. Lastly, double stimulus methodologies are more frequently used.

Interactive variants of the DSIS and ACR methodologies were first proposed by Alexiou et al. [25,46] to assess the quality of geometry-only point clouds in a desktop setting, using the mouse cursor to change the viewpoint. In both studies, Gaussian noise and octree-pruning were employed to simulate position errors from sensor inaccuracies and compression artifacts, respectively. In these user studies, the models were displayed side-by-side using points of minimum size.

The visual quality of colored point clouds was evaluated by Torlig et al. [47] in subjective experiments that were performed in two separate laboratories. Orthographic projections after real-time voxelization of both the reference and the distorted models were shown to the subjects, using the simultaneous DSIS methodology. Point clouds representing both inanimate objects and human figures were selected and compressed using the CWI-PCL codec [49]. The results showed that subjects rate more severely distortions on human models. Moreover, using this codec, marginal gains are brought by color improvements at low geometric resolutions, indicating that the visual quality is rather limited at high sparsity. In a study that followed [28], the same dataset was assessed under the same methodology using a different rendering scheme. The point clouds were rendered using cubes of locally adaptive sizes, with the rating trends being found very similar to those of Torlig et al. [47].

A comprehensive quality assessment study of the MPEG point cloud compression test models is presented with subjective evaluation experiments conducted in two inde-

pendent laboratories [29]. Static, colored point clouds with diverse characteristics were employed and compressed following the MPEG common test conditions. The encoded models were displayed using splats of adaptive size based on local sparsity, and evaluated in an interactive platform using the simultaneous DSIS methodology. As part of the study, subjective experiments under a pairwise comparison protocol were performed to conclude on preferable rate-allocation strategies for geometry, and geometry-plus-color encoding. Based on the findings, human subjects prefer the distortions from regular down-sampling (Octree) over triangulated surface approximations (TriSoup), at both low and high bitrates.

In previous studies, the users were able to rotate, zoom and translate the models, and interact without timing constraints. Yang et al. [48] conducted subjective quality assessment on a large set of widely-employed colored models, allowing only rotation under a fixed distance. Several degradation types affecting both the geometry and the color information were introduced, consisting of octree-pruning, noise injection in the coordinates and the RGB values, random down-sampling, and combinations of the above to further augment the visual impairments. The experiments were conducted using a single stimulus protocol.

Despite the convenience of desktop environments to perform interactive testing, their setup is often less realistic when compared to immersive, XR environments. The first attempt was made by Alexiou et al. [26], making use of an AR setting to evaluate the visual quality of colorless point clouds, subject to octree-pruning and Gaussian noise. A simultaneous DSIS methodology was employed, and a separate session was issued per distortion type. The models were displayed using point primitives of minimum size and were placed as virtual assets in the real world, with users perceiving them via an HMD, and interacting with 6DoF via physical movements.

Perceptual quality of static point clouds in VR was evaluated in a recent study [50]. The users were able to interact with the stimuli with 6DoF via both physical movements and using the controllers, in a virtual scene that was designed to avoid distractions. The color encoding modules of the MPEG G-PCC test model were evaluated using octree-based geometry compression under two double stimulus protocols. The models were displayed using quads of adaptive size that were interpolated before rendering to smooth the surfaces. The user behavior during evaluation was also analyzed to provide further insights. Wu et al. [52] present a study evaluating a large set of colored, static point cloud contents in a 6DoF VR viewing condition. Separate sessions were issued for point clouds depicting human figures and objects. The DSIS methodology with side-by-side inspection was used in all cases, including hidden references to compute DMOS. The point clouds were rendered using minimum point size, while the subjects were able to navigate in the virtual space only by physical movements.

Desktop-based setups were also used for quality assessment of dynamic point clouds. In the work of Mekuria et al. [49], subjective experiments were conducted in the proposed 3D tele-immersive system, where the users were able to interact with naturalistic

(dynamic point cloud) and synthetic (computer generated) models in a virtual scene. The participants were able to navigate in the virtual environment through the use of the mouse cursor in thw desktop setting. The proposed encoding solution (CWI-PCL) that was employed to compress the naturalistic content of the scene was evaluated, among several other aspects of quality (e.g., level of immersiveness and realism).

Visual quality assessment using dynamic point cloud contents in VR under both 3DoF and 6DoF interaction scenarios is presented in the work of Subramanyam et al. [51]. Human figures from real-life acquisition and artificially generated were encoded using the V-PCC and the CWI-PCL, which denotes the anchor codec of the MPEG studies [49]. The models were displayed in the virtual scene using quads of fixed size and evaluated under an ACR-HR protocol. The users were able to navigate by physical movements in the 6DoF scenario, while remaining sited in the 3DoF counterpart. Results showed the superiority of V-PCC at low bitrates, whereas statistical equivalence was found with the MPEG anchor at higher bitrates, depending on the content. Finally, the inability of the codecs to achieve transparent visual quality was remarked. In a subsequent study [53], the subjective quality scores between these 3DoF and 6DoF VR settings were compared to pre-recorded videos visualized on common 2D screens to conclude on the effects of different viewing conditions.

18.1.2.2 User studies for meshes

Many authors have adopted for free interaction in their subjective tests for evaluating the quality of meshes. The majority of these works were performed on 2D screen: Lavoué et al. [54], Corsini et al. [55], and Torkhani et al. [43] conducted subjective experiments based on single stimulus methods (derived from ACR), whereas Lavoué et al. [56] and Silva et al. [57] implemented double stimulus methods (derived from DSIS). In these experiments, the observers were able to freely interact (i.e., free-viewpoint interaction) with the 3D models to evaluate and rate their quality. All those studies considered meshes without color or texture, and they evaluated the impairments introduced by various geometry distortions (e.g., noise, compression, smoothing, watermarking).

An early attempt of a 3D tele-immersive system allowing real-time communication between natural representations of humans and synthetic avatars, was presented by Mekuria et al. in [58]. The natural representations in this setting were rendered as meshes. For purposes of subjective quality evaluation, a pre-recorded natural human moving was employed as the test stimulus. The original representation and three degraded versions after encoding with three real-time mesh coding solutions were subjectively evaluated from near and far distance.

Few experiments involving meshes have been conducted in immersive environments. Christaki et al. [59] subjectively assessed the perceived quality of meshes (without color/texture) subject to different compression codecs in a VR setting using the pairwise comparison method. The content was viewed freely as a combination of natural

navigation (i.e., physical movement in the real-world) and user interaction. Gutiérrez et al. [60] used the dataset of textured meshes provided in [37] to evaluate the perception of geometry and texture distortions in mixed reality (MR) scenarios. They also analyzed the impact of environment lighting conditions on the perceived quality of 3D objects in MR. The experiment was based on the ACR-HR method and the observers were asked to freely explore the displayed 3D models.

18.1.3 Publicly available datasets

Some of the works presented above have publicly released their datasets. Table 18.2 outlines the publicly available subjective quality datasets for 3D content. For meshes, the available datasets concern mostly geometry-only content [19,35,42,43,54,56,57] and are all rather small (see the first 7 rows in Table 18.2). The only public dataset involving meshes with vertex colors is provided by Nehmé et al. [39] and contains 480 distorted stimuli. For textured meshes, three datasets exist: [37], [8] and [41]. The first two datasets contain, respectively, 136 and 28 stimuli, whereas the latter contains more than 343k stimuli, of which 3000 (a generalized and challenging subset) are associated with mean opinion scores (MOS) derived from subjective experiments, and the rest with predicted quality scores (pseudo-MOS), making it the largest quality assessment dataset of textured meshes to date.

Regarding point clouds, only two datasets concern colorless models: [25,26] and [21], with the rest considering colored models. Among the latter, the largest available datasets are the WPC by Su et al. [15], the SJTU-PCQA by Yang et al. [48] and the SIAT-PCQD by Wu et al. [52]. The WPC is composed of point clouds captured in a laboratory setting by the authors, which are degraded by different types of distortions; the SJTU-PCQA makes use of contents that have been extensively utilized in standardization activities under various compression distortions; the SIAT-PCQD involves point clouds from the MPEG and JPEG repositories and a publicly accessible 3D content sharing platform, which are encoded using only V-PCC.

The majority of the reported datasets for both meshes and point clouds were generated through experiments that were conducted on desktop settings. In particular, only the studies presented in [26] and [39,50,52,53] were conducted in immersive environments, with the former performed in AR, and the latter in VR platforms, respectively.

As discussed in Section 18.1, there are various aspects that differ for subjective experiments during the data collection step. All these aspects and different parameters are listed in the columns of Table 18.2. Most of these aspects are either self-describing or introduced at the beginning of this chapter.

Regarding the "Methodology" column, although subjective evaluation methodologies have specific instructions and distinctions among themselves, for the sake of simplicity, we only consider the number of stimuli test participants see to provide a vote. For single stimulus methodologies, the participants decide on the subjective quality by

Table 18.2 Publicly available subjectively annotated datasets for meshes & point clouds.

Dataset	3D Representation	Temporal Variation	Attributes	Mode of Inspection	Methodology	Distortion Types	# Stimuli rated	# Ratings per Stimulus	Raw Scores
LIRIS / EPFL [54]	Mesh	Static	Colorless	Interactive	Single stimulus	Noise addition Smoothing	84	12	✓
LIRIS Masking [56]	Mesh	Static	Colorless	Interactive	Double stimulus	Noise addition	24	11	✓
IEETA Simplification [57]	Mesh	Static	Colorless	Interactive	Double stimulus	Simplification	30	65	X
UWB #1 [35]	Mesh	Static	Colorless	Passive	Pairwise comparison	Compression	63	69	✓
RG-PCD [19]	Mesh	Static	Colorless	Passive	Double stimulus	Octree-pruning	30	126	✓
UWB #2 [42]	Mesh	Dynamic	Colorless	Passive	Multiple stimulus	Compression Noise addition	36	37∼49	X MOS&CI
3D Mesh Animation Quality [43]	Mesh	Dynamic	Colorless	• Passive • Interactive	Single stimulus	Noise addition Compression Transmission error	286	• 16 • 25	✓
LIRIS Textured Mesh [37]	Mesh	Static	Texture maps	Passive (Generated videos)	Pairwise comparison	- On geometry: Compression Simplification Smoothing - On texture: Compression Sub-sampling	• 100×2 renderings • 36×2 renderings	• 11∼15 (Exp.1) • 10∼11 (Exp.2)	X preference matrices
Nehmé et al. [41]	Mesh	Static	Texture maps	Passive (Generated videos)	Double stimulus	Compression Simplification	• 3000 (MOS) • 340750 (Pseudo-MOS)	45	✓

continued on next page

Table 18.2 (*continued*)

Dataset	3D Representation	Temporal Variation	Attributes	Mode of Inspection	Methodology	Distortion Types	# Stimuli rated	# Ratings per Stimulus	Raw Scores
3D Meshes with Vertex Colors [39]	Mesh	Static	Vertex colors	Passive in VR(Slow animations)	Double stimulus	Compression Simplification	480	24	X MOS&CI
G-PCD [25,26]	Point cloud	Static	Colorless	• Interactive • Interactive in 6DoF AR	• Single & Double stimulus • Double stimulus	Noise addition Octree-pruning	50	• 2×20 • 21	✓
M-PCCD [29]	Point cloud	Static	Colored	Interactive	• Double stimulus • 2×Pairwise comparison	Compression	• 240 • 40 & 30	• 40 • 2×25	✓
IRPC [21]	Point cloud	Static	• 2×Colorless • Colored	Passive (Generated videos)	Double stimulus	Compression	• 54 • 54	• 2×20 • 20	X MOS
WPC [15]	Point cloud	Static	Colored	Passive (Generated videos)	Double stimulus	Compression Noise addition Octree-pruning	740	30	X MOS
VsenseVVDB [23]	Point cloud	Dynamic	Colored	Passive (Generated videos)	Double stimulus & Pairwise comparison	Compression	32	19	✓

continued on next page

Table 18.2 (*continued*)

Dataset	3D Representation	Temporal Variation	Attributes	Mode of Inspection	Methodology	Distortion Types	# Stimuli rated	# Ratings per Stimulus	Raw Scores
VsenseVVDB2 [8]	• Point cloud • Mesh	Dynamic	• Colored • Texture maps	Passive (Generated videos)	Single stimulus	Compression	• 136 • 28	23	✓
ICIP2020 [18]	Point cloud	Static	Colored	Passive (Generated videos)	Double stimulus	Compression	96	15∼27	X MOS&CI
PointXR [50]	Point cloud	Static	Colored	Interactive in 6DoF VR	2×Double stimulus	Compression	40	2×20	✓
SJTU-PCQA [48]	Point cloud	Static	Colored	Interactive	Single stimulus	Compression Noise addition Scaling	378	16	X MOS
SIAT-PCQD [52]	Point cloud	Static	Colored	Interactive in 6DoF VR	Double stimulus	Compression	340	38	X DMOS
LB-PCCD [16]	Point cloud	Static	Colored	Passive	Double stimulus	Compression	105	48	✓
2DTV-VR-QoE [53]	Point cloud	Dynamic	Colored	• Interactive in 6DoF VR • Interactive in 3DoF VR • Passive (Generated videos)	Single stimulus	Compression	72	• 26 • 26 • 25	✓

seeing only one stimulus. Similarly, for double stimulus methodologies, participants provide a rating after seeing two stimuli, one of which is generally the reference stimulus (or source model). In multiple stimulus, this number is more than two, and in pairwise comparison, participants choose the better quality stimulus from two stimuli presented to them.

The "# Stimuli Rated" column includes the reference stimuli if they were rated during the test (e.g., as hidden reference). Some datasets might have quality labels in addition to the MOS values that are collected from participants. These quality labels are generally estimated using a quality metric that shows very high correlation to the MOS values. These labels are called pseudo-MOS, and their purpose is to increase the number of labels for metrics training and evaluation.

The "# Ratings per Stimulus" column indicates the number of unique observations made during the experiment, or the number of unique votes that was collected for every stimulus.

Another important parameter for multimedia content quality datasets (images, videos, audios, 3D graphics, etc.) is whether the individual quality scores of each participant were shared and made publicly available. Although all the quality datasets share MOS or preference scores, these values might not be enough to characterize the statistical attributes and distributions of the individual votes [61]. Providing individual votes can allow for further statistical analysis and research into the weaknesses of certain stimuli, use cases, or objective quality metrics [62,63]. Therefore, in Table 18.2, we identify whether the indicated datasets share individual votes from participants in the "Raw Scores" column.

18.1.4 Comparative studies

There are several studies focused on addressing the impact of different aspects (Table 18.1), in subjective quality evaluation of volumetric content. The usage of different types of *3D representation*, the *mode of inspection*, the *display devices*, the *rendering parameters*, and the evaluation *methodologies*, are among the most relevant and popular in the literature. Although some knowledge may be transferred from 2D imaging, which has been well-studied, the effect of different variables for quality assessment of volumetric video can only be quantified through scientific research and experimentation using such contents. Hence, these studies are particularly important, as they can help us better understand and identify interactions between influencing factors.

The first user study aiming at comparing point cloud against mesh *representations* for compression of volumetric video is presented by Zerman et al. in [8]. The Google Draco and JPEG encoding engines were employed for geometry and texture of mesh, respectively, whereas V-PCC and G-PCC were recruited to encode geometry and color of point cloud versions of the contents. As part of the study, the efficiency of the latter MPEG point cloud codecs was also analyzed. All models were evaluated in a passive

protocol using the ACR-HR for both content representations, while point clouds were displayed using fixed-size point primitives. Results show that the point cloud encoding-plus-rendering pipeline leads to better performance at low bitrates, whereas higher quality levels are achieved by the mesh-based counterpart. However, the latter is attained for bitrates that well-exceed the point cloud ones. Finally, among the MPEG alternatives, the superiority of the V-PCC was confirmed.

Similarly, a subjective evaluation of volumetric videos using both point cloud and mesh technologies is detailed in the work of Cao et al. [64]. Several additional factors were considered in the experimental design, among which were the target bitrate, the content resolution, and the viewing distance. To decrease the parameter space, for every target bitrate, a manual identification of the optimal combination for model resolution and compression parameters per viewing distance was performed in a perceptual sense. The selected stimuli were evaluated following passive inspection in two experiments that were carried out. In the first, the subjects rated the visual quality of models that were displayed using both types of content representations under an ACR methodology. In the second, a pairwise comparison between the same models represented as point clouds and meshes was issued. Based on the results, subjects favored the point cloud alternative at lower bitrates. Moreover, the viewing distance was found to be an important factor, and mesh modeling was preferred at closer distances. At higher bitrates and distant inspection, human opinions expressed equal preference.

In the study of Javaheri et al. [21], particular combinations of *representations*, *attributes*, and *rendering* methodologies were examined for quality evaluation of static point clouds, subject to compression distortions. In particular, the experiments were conducted using (a) colorless point clouds, (b) colored point clouds, and (c) colorless meshes, to evaluate the same point cloud distortions. Results show that different scoring behaviors might be observed for the same compression impairments under a different selection. Moreover, the scoring deviations might vary per codec. Finally, it was suggested that texture information might mask underlying geometric distortions.

Regarding the effect of adopting different *modes of inspection* for subjective quality assessment, very few comparisons have been performed. Torkhani et al. [43] performed both passive and interactive experiments for the same dataset of dynamic meshes, using a single stimulus protocol. They concluded that under most kinds of distortions user interaction can affect the perceived quality; however, this impact depends on the nature of the distortion (e.g., global vs local) and is hard to predict. Viola et al. [53] conducted subjective experiments in 3DoF and 6DoF VR as well as with pre-recorded videos in conventional monitors. Two sets of point clouds were employed, subject to compression distortions. For one of the two sets, the viewing condition was deemed to have a significant effect on the distribution of the scores, indicating differences between interactive and non-interactive inspection. For the other set, however, the inspection method had no effect on the scores. The study suggests that conclusions derived from

either a non-interactive or interactive experiment can be roughly generalized, since the effect of the inspection method on the collected scores was, if existent, marginal. In particular, the interaction between codec and inspection method was not significant, meaning that you would draw the same conclusions about the relative performance of one codec with respect to the other in either interactive or non-interactive scenarios. However, this study highlights that other factors, besides visual quality, might be important for the selection of inspection method; for example, the level of presence or immersion, or the discomfort caused by cyber sickness.

Regarding the influence of different *display devices*, the collected user ratings from an AR setting and an interactive desktop setup were compared in [65]. In both experiments, colorless point clouds of minimum size were employed. The results revealed similar rating trends in the presence of Gaussian noise, and differences under octree-pruning. In particular, the authors claim that the former type of degradation leads to clearly perceived artifacts independently of the type of devices, hence leading to high correlation. Differences were also observed with respect to the shape of the models. Finally, higher confidence intervals, associated with subjective scores from the AR setting, suggest a larger number of users to be involved in such experimental setups. Results from a collaborative effort in the framework of the JPEG Pleno on point clouds were reported in [20], where reconstructed mesh models from compressed, colorless point clouds were assessed using various 3D display types/technologies (i.e., passive, active, and auto-stereoscopic) in different laboratories. Inter-laboratory correlations and comparisons with quality scores for the same dataset evaluated in 2D monitors, show very high correlation, suggesting that human judgments are not significantly affected by the display equipment.

The majority of current comparative studies is focused on understanding the impact of employing a different evaluation *methodology* on the obtained quality scores and their accuracy. Specifically, Alexiou et al. [25] compared the results of an ACR and a DSIS test, in which subjects were able to interact with the point clouds viewed on screen. They found that, the DSIS method is more consistent in identifying the level of impairments. The sequential DSIS and a newly proposed variant, namely alternating DSIS, were employed to evaluate point cloud contents subject to color compression distortions in [50]. In the former protocol, the reference model is presented to the user followed by the distorted, whereas in the latter, the user is allowed to toggle between the reference and the distorted at will. The experiments were conducted in VR with users interacting with 6DoF. The results indicated that the alternating DSIS protocol leads to lower uncertainty for the perceived distortions; it is faster, and generally preferred by the participants. Recently, a comprehensive study [40] compared the performance of three of the most prominent subjective methodologies, with and without explicit references, namely ACR-HR, DSIS and SAMVIQ, to determine the best one for evaluating the perceived quality of 3D graphics, especially in VR. The study was conducted in a VR

environment using a dataset of meshes with vertex colors animated with slow rotation. Results assert that the presence of an explicit reference is necessary to improve the accuracy and the stability of the method. DSIS tends to be the most suitable method, in terms of accuracy and time-efficiency, to assess the quality of 3D graphics. Authors recommended the use of at least 24 observers for DSIS tests.

18.2. Objective quality assessment

Although subjective quality assessment provides ground truth quality scores for visual stimuli, it is not feasible to carry out user studies for each scenario, especially when there is a need to determine the quality for a large number of contents or for real-time applications. In these cases, objective quality assessment methods (or objective quality metrics) are particularly useful, enabling algorithmic quality estimation using mathematical calculations and signal processing approaches.

Simple geometric or color distances/errors between 3D models are weakly correlated with human perception, since they ignore perceptual characteristics of the HVS [23,54,66], analogously to the mean square error (MSE) and peak signal-to-noise ratio (PSNR) measurements in 2D imaging. Therefore current efforts are concentrated on perceptually driven visual quality metrics, which can be primarily distinguished in top-down and bottom-up approaches. The former treat the HVS as a black box and capture modifications in content features that are induced by distortions to estimate perceived quality. The latter rely on computational models that describe properties of the HVS, mainly to determine the visibility of errors caused by distortions. With the rise of machine learning, a third category has recently emerged, consisting of metrics that rely on purely data-driven approaches, which do not demand any explicit model.

Independently of the design, objective quality metrics can be categorized based on their requirement for the original content (i.e., reference) at execution time as full-reference (FR), reduced-reference (RR), and no-reference (NR). For FR metrics, the distorted model is compared to its reference. For RR metrics, some reference data are required as inputs, whereas for NR, no reference information is necessary. FR metrics are generally employed to drive lossy processing operations, such as compression, transmission, simplification, and watermarking. However, they have higher computational overhead and are not always applicable, as the original content is not always available. RR metrics make use of lightweight, descriptive features that are extracted from both the reference and the distorted contents for comparison purposes. They are typically employed when it is inefficient, or impractical to provide the entire original content, such as in a video streaming scenario. NR metrics are the most practical in terms of usage, yet they are often rather limited in terms of scope, i.e., tuned for a particular type of distortion with limited generalization capabilities.

Finally, considering their operating principle, 3D quality metrics can be classified as model-based and image-based (also known as rendering-based or projection-based)

metrics. The model-based metrics operate on the 3D model itself (either mesh or point cloud) and its attributes, such as texture maps or color values. The image-based metrics function on the image domain and usually on projected views of 3D models on 2D planar arrangements; i.e., they often apply image quality metrics (IQMs) on 2D snapshots of the rendered 3D model. The rest of this section is structured according to this classification: In Section 18.2.1, model-based methods for point clouds and meshes are described, whereas in Section 18.2.2, image-based metrics are reported for both types of content representations.

18.2.1 Model-based quality metrics

The majority of model-based quality metrics are based on top-down FR approaches. In FR quality metrics, a correspondence function is essential to enable comparisons between the original (or reference) and the distorted contents. In conventional 2D imaging, this is easily achieved by matching the pixel grids of the reference and the distorted contents. However, this is not the case for 3D data, whose topology is altered by geometric distortions that typically introduce dislocation and/or removal of 3D points. Given the different 3D point (or vertex) populations and coordinates, the point matching (i.e., correspondence) step between the reference and the distorted models becomes an ill-posed problem. After establishing a correspondence, errors between attributes or features associated to the matched points are computed.

Often, 3D data quality metrics identify point matches using the nearest neighbor algorithm for simplicity reasons. In particular, following the most common conventions, the original model is selected as the reference $\mathcal{R}$ and the distorted model is set under evaluation $\mathcal{T}$. Considering point clouds (or meshes), for every point (or vertex) $\mathbf{t} \in \mathcal{T}$, the nearest reference point (or vertex) $\mathbf{r_t} \in \mathcal{R}$ is identified, and a local error $e(\phi_{\mathbf{t}}, \phi_{\mathbf{r_t}})$ is computed between corresponding features $\phi_{\mathbf{t}}$ and $\phi_{\mathbf{r_t}}$. A global quality score $q_{\mathcal{T}\to\mathcal{R}}$ is obtained by pooling the local errors, as given in Eq. (18.1):

$$q_{\mathcal{T}\to\mathcal{R}} = \frac{1}{|\mathcal{T}|}\left(\sum_{\mathbf{t}\in\mathcal{T}} e(\phi_{\mathbf{t}}, \phi_{\mathbf{r_t}})^m\right)^{1/n}, \tag{18.1}$$

where $|\mathcal{T}|$ is the number of points (or vertices) of $\mathcal{T}$ and $m, n \geq 1$. Using $n = 1$, and $m = 1$ or $m = 2$, the average or the MSE are obtained, respectively.

Note that $q_{\mathcal{T}\to\mathcal{R}}$ is an asymmetric measurement, as $q_{\mathcal{T}\to\mathcal{R}} \neq q_{\mathcal{R}\to\mathcal{T}}$. That is, by selecting the distorted model as the reference, different sets of matched points (or vertices) are obtained, resulting in different global quality scores. To obtain quality predictions that are independent of the reference selection, it is common to use both models as reference and apply a symmetric operation $f(\cdot)$ on the exported global quality scores, such as the average or the maximum, as shown in Eq. (18.2).

$$q = f\left(q_{\mathcal{T}\to\mathcal{R}}, q_{\mathcal{R}\to\mathcal{T}}\right). \tag{18.2}$$

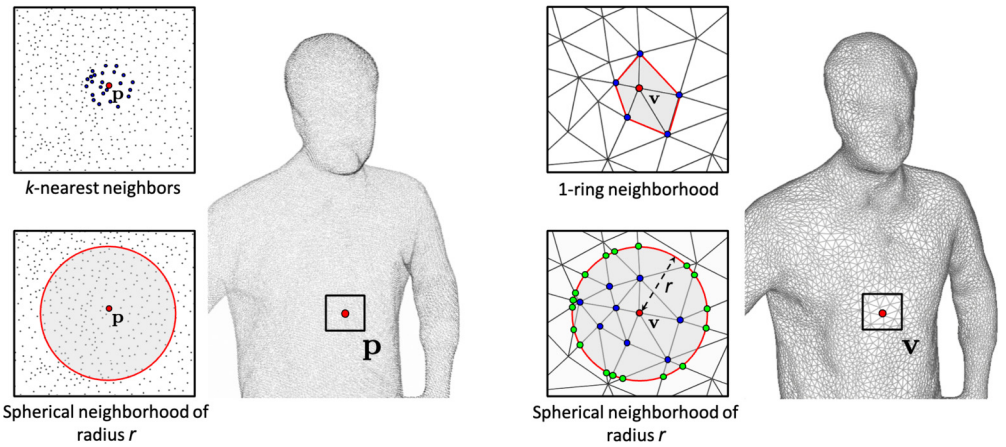

Figure 18.4 Local neighborhood computation around a point **p** in a point cloud (left) and a vertex **v** in a mesh (right). For the point cloud, the k-nearest neighbors with $k = 24$ and the range search with radius r (r-search) approaches are depicted. The 1-ring neighborhood as well as the r-search are illustrated for the mesh.

Another approach that has been widely used to provide a global quality score, is the Hausdorff distance. It is defined as the greatest out of all the distances between the points (or vertices) of $\mathcal{T}$ and the nearest/corresponding points (or vertices) of $\mathcal{R}$. It can be derived by using max pooling (instead of the aggregation in Eq. (18.1)) of local errors that measure the Euclidean distance between a point (or vertex) $\mathbf{t}$ and its nearest $\mathbf{r_t}$. The Hausdorff distance of the point cloud (or mesh) under evaluation $\mathcal{T}$ from the reference $\mathcal{R}$ is computed as follows:

$$q_{\mathcal{T}\to\mathcal{R}} = \max_{\mathbf{t}\in\mathcal{T}}\{\min_{\mathbf{r}\in\mathcal{R}}\{d(\mathbf{t},\mathbf{r})\}\} = \max_{\mathbf{t}\in\mathcal{T}}\{d(\mathbf{t},\mathbf{r_t})\}, \tag{18.3}$$

where $d(\cdot)$ is the Euclidean distance.

Finally, it is rather frequent for both point cloud and mesh model-based metrics to take into consideration local neighborhoods around a queried point (or vertex) to compute an attribute or a feature ϕ. For point clouds, the most common algorithms are the k-nearest neighbor and the range search with radius r, denoted as k-nn and r-search, respectively, and shown on the left side of Fig. 18.4. The former identifies the nearest k points to a queried point $\mathbf{p}$, whereas the latter returns all points enclosed in a sphere with center $\mathbf{p}$ and radius r. For meshes, the 1-ring neighborhood and the r-search, illustrated on the right side of Fig. 18.4, are employed. The first refers to the set of all vertices connected with the queried vertex $\mathbf{v}$ by an edge. The second, is defined as the connected set of vertices belonging to the sphere with center $\mathbf{v}$ and radius r. In this case, the intersections between this sphere and the edges of the mesh are also added

to the neighborhood. Note that the k-nn (for point clouds) and 1-ring (for meshes) approaches lead to neighborhoods of arbitrary extent, depending on the point density or vertex sampling of the model. The k-nn approach has a fixed population (equal to k), and the 1-ring is straightforward to compute. Concurrently, the r-search approach identifies same volumes that may enclose varying number of samples.

In what follows, we first discuss the model-based metrics for point clouds, and then those for meshes.

18.2.1.1 For point clouds

The development of point cloud objective quality metrics has been an active research field the last five years. This interest was fueled by the MPEG and JPEG standardization activities on point cloud compression [2,3], which required reliable solutions for quality assessment of point cloud compression distortions.

Early developments of model-based predictors employed simple distances between attributes of matched points to measure local errors, as shown in Fig. 18.5. The point-to-point metric denotes the earliest attempt, with the geometric variant computing the Euclidean distance between point coordinates to measure the geometric displacement of distorted samples from their reference positions [67]. Setting $n = 1$ and $m = 2$ in Eq. (18.1), the point-to-point metric with MSE is computed, also known as D1 [68]:

$$\mathrm{D1} = \frac{1}{|\mathcal{T}|} \sum_{\mathbf{t} \in \mathcal{T}} d(\mathbf{t}, \mathbf{r}_\mathbf{t})^2, \tag{18.4}$$

where $d(\cdot)$ is the Euclidean distance. Analogously, the point-to-point variant for color distortions measures the error between RGB color values or YUV intensities of matched points, effectively simulating the MSE that has been widely used for 2D images [69], while the PSNR version of this measurement is obtained straightforwardly..

The point-to-point metrics have low complexity; however, they do not account for perceptual characteristics of the HVS. An early alternative to capture geometric distortions based on distances that are more perceptually relevant, is the point-to-plane metric [67]. This method relies on the projected error of distorted points across reference normal vectors. Thus local errors measure the deviation of distorted points from linearly approximated reference surfaces. A global degradation score is typically obtained using the MSE, as given below following the conventions of Fig. 18.5, which is also referred to as D2 [68]:

$$\mathrm{D2} = \frac{1}{|\mathcal{T}|} \sum_{\mathbf{t} \in \mathcal{T}} |\vec{u} \cdot \vec{n}_{\mathbf{r}_\mathbf{t}}|^2. \tag{18.5}$$

Beyond MSE, the Hausdorff distance (Eq. (18.3)) has been additionally used with both point-to-point and point-to-plane metrics. Finally, the geometric PSNR has been proposed for both metrics to account for differently scaled contents [70], using either

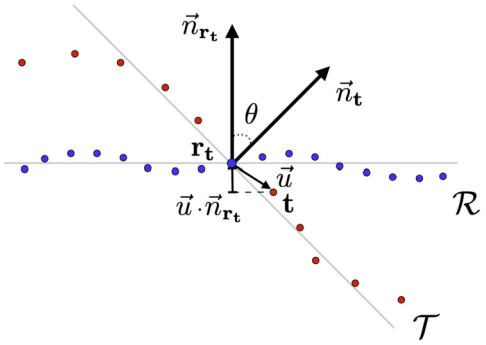

Figure 18.5 Error quantification between a distorted point $\mathbf{t}$ and a reference point $\mathbf{r_t}$ from simple point-wise distances. Setting $\vec{u} = \mathbf{t} - \mathbf{r_t}$ and θ the angle between corresponding normal vectors, the local error for point-to-point is $d(\mathbf{t}, \mathbf{r_t}) = \|\vec{u}\|^2$, for point-to-plane is $|\vec{u} \cdot \vec{n}_{\mathbf{r_t}}|$, and for plane-to-plane is $1 - 2\min\{\theta, \pi - \theta\}/\pi$.

the voxel grid's diagonal or the maximum nearest-neighbor distance of the original content as the peak value.

The above metrics were employed in the MPEG standardization activities since the beginning. As a consequence, they were widely used and tested by the research community, attracting interest and inspiring submissions to improve their performance. For instance, the generalized Hausdorff distance was proposed to mitigate the sensitivity of the Hausdorff distance in outlying points by excluding a percentage of the largest individual errors [71]. A revised geometric PSNR calculation was proposed in [72], setting as peak value the average over distances between neighbors in 3D space, or after projection onto local planes, to represent the intrinsic resolution or rendering resolution of the content, respectively.

Another early-developed method evaluating geometry-only distortions is the plane-to-plane metric, proposed by Alexiou et al. [73]. This metric estimates the difference in orientation between local surface approximations of the original and the distorted point clouds. This is achieved by computing the angular similarity between unoriented normal vectors from locally fitted surfaces. In particular, the angle between the two normal vectors is computed as follows:

$$\theta = \arccos\left(\frac{\vec{n}_{\mathbf{r_t}} \cdot \vec{n}_{\mathbf{t}}}{\|\vec{n}_{\mathbf{r_t}}\| \, \|\vec{n}_{\mathbf{t}}\|}\right), \tag{18.6}$$

where the angular similarity is given as

$$\text{Angular similarity} = 1 - \frac{2\min\{\theta, \pi - \theta\}}{\pi}. \tag{18.7}$$

The plane-to-plane metric relies on the computation of normal vectors and its performance is affected by how they approximate the underlying surfaces. An insightful analysis for the calibration of this metric is provided in [74].

In the same list of geometry-only model-based quality metrics lies the point-to-distribution, introduced by Javaheri et al. [75], which computes the Mahalanobis distance between a point and a reference neighborhood. In this case, the geometric deviation is measured with respect to the distribution of reference samples, thus accounting for the local reference topology. This metric was lately extended to capture color degradations by applying the same formula to luminance attributes [76]. The two quality scores obtained for geometry and texture were simply averaged to provide a final predicted quality score.

More recent proposals rely not only on surface properties extracted from point samples, but also on the utilization of statistics to capture relations between points that lay in the same local neighborhood. For that purpose, a correspondence is established between points in the point cloud under evaluation $\mathcal{T}$, and the relative reference point cloud $\mathcal{R}$. Then, statistics are computed based on the neighborhood surrounding the points, as seen in Fig. 18.6. An initial metric towards this direction is PC-MSDM by Meynet et al. [77], which is based on the relative difference between local curvature statistics (mean, standard deviation and covariance of curvature). The PC-MSDM was later extended to colored point clouds by incorporating local statistical measurements of luminance, chrominance, and hue components to evaluate textural impairments. A proposed weighting function regularizes the contributions of each feature in the final quality prediction. The new metric is called PCQM [78]. Both PC-MSDM and PCQM, instead of using nearest neighbors, employ the reference points and their projections onto the quadric surfaces fitted to the distorted model as correspondences.

Alexiou et al. proposed PointSSIM [79], which relies on a similar logic, capturing perceptual degradations based on the relative difference of statistical dispersion estimators applied on local populations of location, normal, curvature, and luminance data. An optional pre-processing step of voxelization is proposed to enable different scaling effects and reduce intrinsic geometric resolution differences across contents. The VQA-CPC metric, by Hua et al. [80] depends on statistics of geometric and color quantities. These quantities are obtained by computing the Euclidean distance between every sample from the arithmetic mean of the point cloud, considering geometric coordinates and color values, respectively. The color point cloud metric based on geometric segmentation and color transformation (CPC-GSCT) denotes an extension, involving a partition stage of the point cloud, before the extraction of features per region [81]. The geometric features consist of statistical moments applied on Euclidean distances, angular distortions, and local densities, which are weighted according to the roughness of a region. The textural features rely on the same statistics after conversion to the HSV color model.

More recently, the PointPCA metric was presented by Alexiou et al. [82], making use of statistics applied on a series of geometric and textural descriptors. The former are

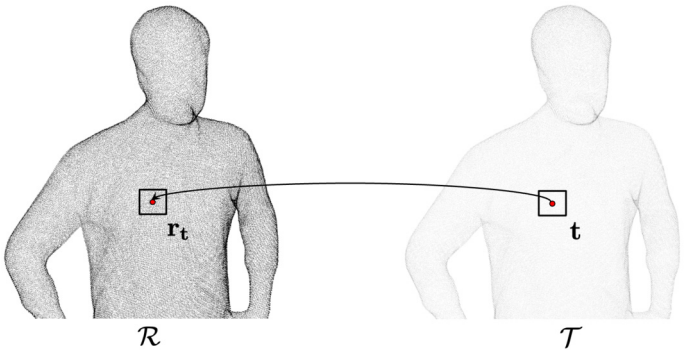

Figure 18.6 Computation of statistics considering a neighborhood around every point. Points that belong to the point cloud under evaluation $\mathcal{T}$ are matched with points from the reference $\mathcal{R}$, and corresponding statistics are compared to obtain local errors. Normally, a global quality score is computed by pooling, per Eq. (18.1).

extracted from measurements obtained after performing principal component analysis (PCA) in support regions defined around point samples, while the latter consist of luminance intensities. The proposed geometric descriptors capture the dispersion of points distribution, the dimensionality, variation, and roughness of the underlying surface, as well as the parallelity with respect to the coordinate system axes. An individual prediction is obtained from every geometric and textural descriptor, with a final quality score obtained via a weighted average.

The GraphSIM by Yang et al. [83] denotes a graph signal processing-based approach, which evaluates statistical moments of color gradients computed over graphs. The graphs are constructed around keypoints of the reference content, and are identified after high-pass filtering on its topology. A more recent, multi-scale version of this metric, namely MS-GraphSIM, is presented by Zhang Y. et al. [84] The proposed multi-scale point cloud representation is achieved by low-pass filtering on color, point down-sampling, and region shrinking. The FQM-GC by Zhang K. et al. [85] is another solution that extracts geometry features from graphs constructed per partition after geometry segmentation, with the graph signal considering normal vectors. Moreover, a color segmentation step is employed and a colorfulness index is computed to weight the obtained segments accordingly. Color histograms and the relative difference of several moments of chrominance components are employed to estimate color distortions per segment.

Xu et al. [86] present the EPES, a point cloud quality metric based on potential energy. In this method, a number of points are selected, called origins, after applying a high-pass filtering operation in the topology of a point cloud. Local neighborhoods are formed around these origins, and the potential elastic energy needed to move the enclosed points from the origin to their current state (considering both geometry and

color) is computed. A global score is obtained by aggregating the individual elastic potential energies across origins. Additionally, a local score is obtained as the cosine similarity between the direction of forces needed to transfer a reference and a corresponding distorted point from their origin to their current positions. Global and local features are pooled together to provide a final quality score.

In the work of Diniz et al. [87], local binary patterns on the luminance channel are applied in local neighborhoods. This work is later extended [88] to additionally take into consideration the point-to-plane distance between point clouds, and the point-to-point distance between corresponding feature maps in the quality prediction. A variant descriptor, namely, local luminance patterns, is proposed in [89]. This work also introduces a voxelization stage in the metric's pipeline to alleviate its sensitivity to different voxelization parameters. In [90], a texture descriptor to compare neighboring color values using the CIEDE2000 distance is proposed. The color differences are coded as bit-based labels, which denote frequency values of predefined difference intervals. An extension is presented in [91], namely, BitDance, which incorporates a geometric descriptor that relies on the comparison of neighboring normal vectors, resulting in bit-based labels similarly to the texture counterpart.

A set of texture-only metrics has been proposed by Viola et al. [92], which relies on histograms or correlograms of luminance and chrominance components to characterize the color distributions of distorted and reference point cloud data. A global quality score is obtained by weighted combination of the proposed color-based predictor and the point-to-plane metric.

The previous works refer to FR quality metrics. Regarding RR approaches, the first attempt was reported by Viola et al. in [93]. The algorithm is based on global features extracted from the location, luminance, and normal data, with a weighted average used to combine the distortions into a single quality score. More recently, an RR metric for point clouds encoded with V-PCC was presented by Liu et al. [94]. The prediction is a linear model of geometry and color quantization parameters, with parameters determined by a local and a global color fluctuation feature that accounts for the different impact of compression artifacts on contents.

An NR method was recently proposed by Hua et al. in [95], namely, BQE-CVP. This method relies on geometric features based on point distances, normals, curvatures and point density, which are estimated after segmentation and weighted according to local roughness, similarly to the same authors' previous work in [81]. Texture degradations are computed as statistical moments of distortion maps obtained after applying the just noticeable distortion [96] on point cloud projections. Moreover, features based on gray-texture variations, color entropy, and color contrast, are extracted. Finally, a joint feature based on a geometric-color co-occurrence matrix is proposed.

The performance of well-established, pre-trained convolutional neural network (CNN) architectures for classification, was investigated to assess the quality of the point

clouds, after necessary adjustments by Chetouani et al. in [97]. Geometric distance, mean curvature, and luminance values are employed and form patches. A patch quality index is computed using a CNN model, and a global quality index is obtained after pooling. Quach et al. [98] extend the use of perceptual loss from 2D images to point clouds, which are represented as voxel grids with binary occupancy or truncated distance fields. The perceptual loss is applied on the latent space, after passing through a simple auto-encoding architecture that is composed of convolution layers.

18.2.1.2 For meshes

Many mesh visual quality metrics have been proposed in the literature in the past 15 years [99,100]. Existing metrics are mostly FR. Most of them also follow the classical top-down approach used in image quality assessment: local feature differences between the reference and distorted meshes are computed at vertex level, and then pooled over the entire 3D model to obtain a global quality score. Whereas pioneering techniques were limited to evaluating geometry distortions only, most recent ones incorporate color/texture information. Moreover, machine learning and, more recently, deep learning approaches are gaining in popularity. This allows, among other benefits, the emergence of NR methods. The paragraphs that follow detail existing mesh quality metrics.

As mentioned earlier, pioneering metrics evaluated only geometric distortions, i.e., they rely on geometric characteristics of the mesh without considering its appearance attributes. Primary works used simple geometric measures, such as Hausdorff distance [101], MSE, root mean squared (RMS) error [102] and PSNR. These measures quickly demonstrated a poor correlation with the human vision, since they ignore perceptual information [103]. Hence, encouraging the development of more perceptually driven visual quality metrics. One of the first proposed metrics combined the RMS geometric distance between corresponding vertices with the RMS distance of their Laplacian coordinates, which reflect the degree of surface smoothness [104]. A strain energy field-based measure (SEF) was also developed by Bian et al. in [105]. This metric is based on the energy introduced by a specific mesh distortion; that is, the more the mesh is deformed, the greater the probability of perceiving the difference between the reference and distorted meshes. Some authors were inspired by IQMs. For instance, Lavoué et al. [54,66] proposed two metrics, called mesh structural distortion measure (MSDM) and MSDM2, inspired by the well-known SSIM [106]. In particular, the authors extended the SSIM to meshes by using the mesh curvature as an alternative for the pixel intensities. MSDM2 is adapted for meshes with different connectivities. Torkhani et al. [107] also proposed a metric based on local differences in curvature statistics. They included a visual masking model to their metric. Other works considered the dihedral angle differences between the compared meshes to devise their metric, such as the dihedral angle mesh error (DAME) metric [35]. The above metrics consider local variations

at the vertices or edges. Corsini et al. [55] proceeded differently. They computed one global roughness value per 3D model considering dihedral angles and variance of the geometric Laplacian, and then derived a simple global roughness difference. In a similar approach, Wang et al. [108] proposed a metric called fast mesh perceptual distance (FMPD) based on global roughness computed using the Gaussian curvature. A survey [99] detailed these works and showed that MSDM2 [66], DAME [35], and FMPD [108] are excellent predictors of visual quality.

Besides these works on global visual quality assessment (top-down approaches adapted for supra-threshold distortions), few works based on bottom-up approaches were proposed. Nader et al. [109] introduced a bottom-up visibility threshold predictor for 3D meshes. Guo et al. [110] also studied the local visibility of geometric artifacts and showed that curvature could be a good predictor of distortion visibility.

Several works used machine learning techniques in assessing the quality of meshes, with multi-linear regression adopted to optimize the weights of several mesh descriptors [111], or support vector regression (SVR) used to fuse selected features to obtain a quality metric [112]. Recently, a machine learning-based approach for evaluating the quality of 3D meshes was proposed, in which crowd-sourced data is used, while learning the parameters of a distance metric [113].

Moving to dynamic meshes, Váša et al. [42] proposed a metric, called STED, based on the comparison of mesh edge lengths and vertex displacements between two animations. Torkhani et al. [43] devised a quality metric for dynamic meshes, which is a combination of spatial and temporal features. In more recent work, Yildiz et al. [114] developed a bottom-up approach incorporating both the spatial and temporal sensitivity of the HVS to predict the visibility of local distortions on the mesh surface.

For some use cases, the reference might not be available. Therefore NR quality assessment metrics are needed. Unlike FR metrics, few NR quality metrics for meshes have been proposed in the literature. These metrics are based on data-driven approaches (machine learning). Abouelaziz et al. [115] proposed an NR metric that relies on the mean curvature features and the general regression neural network (GRNN) for quality prediction. The blind mesh quality assessment index (BMQI), proposed in [116], is based on the visual saliency and SVR, whereas that proposed in [117] is based on dihedral angles and SVR. Abouelaziz et al. [118] also used CNNs to assess the quality of meshes. The CNN was fed with perceptual hand-crafted features (dihedral angles) extracted from the mesh and presented as 2D patches.

All the works presented above consider only the geometry of the mesh, and therefore only evaluate geometric distortions. Regarding 3D content with color or material information, little work has been published. For meshes with diffuse texture, Pan et al. [36] derived from the results of a subjective experiment a quantitative metric that approximates perceptual quality based on texture and geometry (wireframe) resolution. Tian et al. [119] and Guo et al. [37] proposed metrics based on a weighted combination

of a global distance on the geometry and a global distance on the texture image. Tian et al. [119] combined the MSE computed on the mesh vertices with that computed on the texture pixels, whereas Guo et al. [37] linearly combined MSDM2 [66] (for mesh quality) and SSIM [106] (for texture quality) metrics. These metrics combine errors computed on different domains (mesh and texture image). Very recently, Nehmé et al. [39] introduced the color mesh distortion measure (CMDM), which, to this date, is the only model-based quality metric for meshes with colors attributes that works entirely on the mesh domain. This metric incorporates perceptually relevant geometry and color features and is based on a data-driven approach. It can be viewed as the mesh version of the point cloud metric PCQM [78].

As can be seen, most existing model-based quality metrics ignore the visual saliency information, yet finding salient regions (regions that attract the attention of observers) has become a useful tool for many applications, such as mesh simplification [120] and segmentation [121], and quality control of VR videos (360 videos) [122,123]. A recent work has investigated how incorporating saliency information into a model-based metric can improve the predicted quality [124]. Authors devised an extension of the CMDM metric [39] by combining its geometry and color features with the visual attention complexity (VAC) feature based on visual saliency dispersion proposed in [125]. Integrating the VAC was found to improve the overall performance of CMDM, especially when assessing the quality of geometrically quantized stimuli.

18.2.2 Image-based approaches

To evaluate the quality of 3D content, several authors considered IQMs computed on rendered snapshots, as depicted in Fig. 18.7. These approaches can be efficient since the field of image quality assessment is highly developed, and many successful IQMs have been introduced, such as the Sarnoff VDM [126], SSIM [106] (and its derivatives), VIF [127], FSIM [128], HDR-VDP2 [129], iCID [130], BLIINDS [131], GMSD [132], DeepSIM [133], LPIPS [134], WaDIQaM [135], NIMA [136], and PieAPP [137].

The image-based approach was first used to drive perceptually based tasks, such as mesh simplification [138,139]. So far, mainly FR approaches have been proposed for quality evaluation of 3D data of both meshes and point clouds. That is, views of the original and the distorted contents are captured under identical camera parameters, and a quality prediction is obtained as an average, or a weighted average of individual objective scores.

Image-based metrics allow holistic capture of both topology and color distortions as reflected by the corresponding rendering application. However, several factors affect the image-based metrics' computations. In particular, the rendering scheme that is employed to display the 3D data together with the environmental and lighting conditions, the number of cameras (or viewpoints), the configuration of each camera's parameters for the acquisition of model views, and the pooling of quality scores obtained for differ-

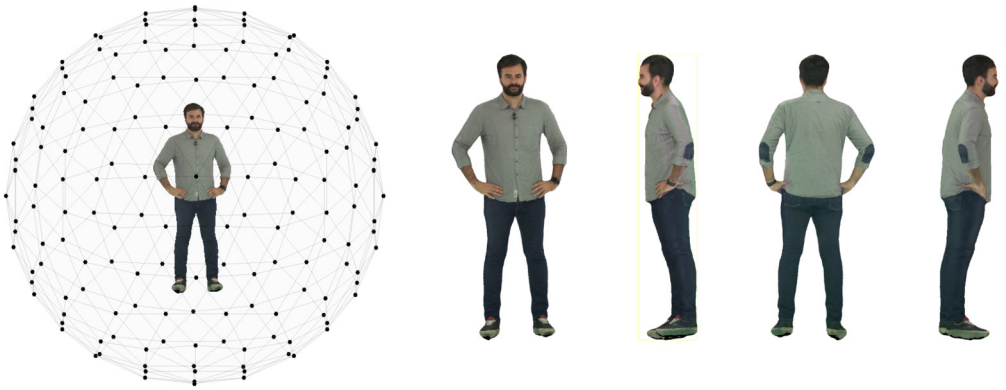

Figure 18.7 A common camera arrangement to capture views of a 3D model is illustrated on the left. The camera positions (i.e., black points) are typically selected to be uniformly distributed across a surrounding sphere. Snapshots of the 3D model are presented on the right, as captured from the front, right, back, and left cameras.

ent views into a single global quality score. Hence, image-based metrics are considered to be rendering-dependent and view-dependent solutions [29,54], which are opposed to the rendering-agnostic, model-based methods. The impact of such factors is further discussed in subsection 18.2.3.

In what follow, we first present the image-based metrics for point clouds, then those for meshes.

18.2.2.1 For point clouds

Image-based approaches were first used for point cloud imaging in the work of de Queiroz et al. [140]. Their prediction accuracy on point cloud contents was initially examined by Torlig et al. [47]. Concretely, the PSNR, SSIM [106], MS-SSIM [141], and VIF [127] (applied on the pixel domain) were executed on images after orthographic projection of the reference and distorted point clouds on the faces of a surrounding cube. The results showed that the MS-SSIM was the best candidate, achieving better performance than the model-based alternatives available at the time. The same conclusions regarding the effectiveness of MS-SSIM were drawn in another study using the same point cloud contents, but under a different rendering technique; that is, the point clouds were rendered using cubic primitives of locally adaptive size both for subjective evaluation and for the computation of image-based metrics; [28].

The first image-based metric tailored for point cloud contents was proposed by Yang et al. [48], relying on a weighted combination of global and local features extracted from texture and depth images. Specifically, the Jensen-Shannon divergence on the luminance channel serves as the global feature, whereas a depth edge map that re-

flects discontinuities, a texture similarity applied on color components, and an estimated content complexity factor account for the local features. Another approach, proposed by He et al. [142] was to project color and curvature values on planar surfaces. In this case, color impairments are evaluated using probabilities of local intensity differences, together with statistics of their residual intensities, and similarity values between chromatic components. The geometric distortions are evaluated based on statistics of curvature residuals.

A hybrid method that uses both image- and model-based algorithms is presented by Chen et al. [143]. In particular, the point clouds are divided into non-overlapping partitions, called layers. A planarization process takes place at each layer, before applying the IW-SSIM [144] to assess geometric distortions. Color impairments are evaluated using RGB-based variants of similarity measurements defined in [78]. A linear model is employed to assign optimal weights on the defined features.

A deep neural network architecture, namely PQA-Net, was proposed by Liu et al. [145] for NR quality assessment of point clouds. In this method, features are extracted from multiple views after a series of CNN blocks and, after fusion, they are shared between a distortion identifier and a quality predictor to obtain a final quality score.

Wu et al. [52] apply popular IQMs on patches from geometry and texture images. The patches are obtained after segmenting the reference point cloud into point clusters based on normal vectors. To ensure pixel matching between the reference and the distorted patch, for every reference point, its nearest distorted point is identified, and both are projected on the same pixel locations. Results show substantial improvements with respect to the application of the same IQMs on the six sides of models' bounding boxes. The IW-SSIM was found to achieve best performance.

A learning-based approach based on patches from projected maps of geometry and texture, as implemented in V-PCC, is presented by Tao et al. [146]. The proposed network makes use of a joint color-geometric feature extractor, two-stage multi-scale feature fusion, and spatial pooling. The extractor is composed of sequential CNNs to extract multi-scale features from geometry and color patches separately, with corresponding features maps subsequently fused. The spatial pooling module consists of two fully-connected layer branches that perform (a) quality prediction and (b) weight allocation, per patch. The final score is obtained as a weighted average across all patches.

18.2.2.2 For meshes

IQMs, notably VDP [147] and SSIM [106], were used to study the relationship between the viewing distance and the perceptibility of model details to optimize the level-of-detail (LoD) design of complex 3D building facades [148]. SSIM [106] was also used to optimize textured mesh transmission [149]. Considering a view-independent approach, the RMS error was computed on snapshots taken from different viewpoints (different

camera positions regularly sampled on a bounding sphere) to evaluate the impact of simplification on 3D models [150].

Recently, several authors have started to exploit CNNs to assess the quality of meshes using an image-based approach. Most of the existing works considered geometry-only meshes (without color attributes). In [151], the CNN was fed with 2D rendered images of the mesh generated by rotating the object. Another quality metric for meshes was devised by extracting feature vectors from 3 different CNN models and combining them using an extension of the compact bi-linear pooling (CMP) [152]. The authors used a patch-selection strategy based on mesh saliency to give more importance to perceptually relevant (attractive) regions. In fact, not all regions of the 3D model image receive the same level of attention from observers.

A more recent metric called graphics-LPIPS was proposed for assessing the quality of rendered 3D graphics [41]. The metric is an extension of the LPIPS metric (originally designed for images and perceptual similarity tasks) [134], which has been adapted to 3D graphics and quality assessment tasks based on DSIS. Graphics-LPIPS is computed on patches of snapshots of the rendered 3D models and employs a CNN (the AlexNet architecture more precisely) with learning linear weights on top. The overall quality of the 3D model is derived by averaging local patch qualities.

18.2.3 Comparison between model-based and image-based approaches

Several works compared the performance of image-based metrics and model-based approaches for quality assessment of 3D models [16–18,28,29,39,47,52,103]. Results indicate that both approaches are having their merits, with model-based generally showing higher generalization capabilities across contents and distortions. Table 18.3 summarizes advantages and disadvantages, as well as use cases of each approach.

The main advantage of using image-based metrics to evaluate the visual quality of 3D objects is their natural handling of complex interactions between different data properties involved in the appearance (geometry, color or texture information, and normals), which avoids the problem of how to combine and weight them [150]. For instance, using IQMs on projected views of 3D models simultaneously captures geometric and chromatic degradation as reflected in the renderer, in addition to the natural incorporation of the complex rendering pipeline (computation of light material interactions and rasterization), thus capturing 3D content as experienced/perceived by users.

On the other hand, these methods require prior knowledge of the final rendering of the stimuli, i.e., the lighting conditions and the viewpoint, since they operate on 2D rendered snapshots. Additionally, they depend on the choice of 2D views employed to estimate a quality score. In particular, the selection of camera positions, camera parameters, number of viewpoints, and pooling applied across different views, will lead to different quality characterizations for the same 3D model.

Table 18.3 Overview of the advantages, disadvantages, and use cases of the image-based and model-based approaches.

	Model-based approaches	Image-based approaches
Advantages	• Independent of the final rendering & the displayed viewpoint • Practical for driving processing operations	Natural ability to handle: • multimodal nature of data • complex rendering pipeline
Disadvantages	Sophisticated algorithms to handle: • multimodal nature of data • complex rendering pipeline	• Prior knowledge of the final rendering & the displayed viewpoint • The choice of 2D views, their number, and the pooling method
Use cases	• Evaluating different distortions applied to different 3D models • Driving perceptually based tasks	• Evaluating the quality of different versions of the same object under a single type of distortion

In these frameworks, the number of views and the camera parameters are set to cover the maximum surface of a model under evaluation. However, using a large number of views/cameras leads to redundancies and extra computational costs, without guaranteeing performance improvements, as indicated in [28]. When applying IQMs on projected views of a 3D model, excluding pixels that don't belong to the effective part of the displayed model (i.e., background filtering), was found to improve the accuracy of the predicted quality [28]. Moreover, non-uniform weightings that increase the impact of quality scores from views that are more relevant may improve the prediction performance. Alexiou et al. [28] showed that estimating the global quality score by incorporating importance weights based on user inspection time is beneficial in terms of prediction accuracy (i.e., better performance than uniform weighting) and computational costs (i.e., less views are required to be captured, especially in dense camera arrangements). Wu et al. [52] incorporated a weighting function based on the ratio of projected area of that model view with respect to the total amount, observing performance improvements. Weighted views have been also considered in [103] for objective quality evaluation of meshes, with importance weights obtained based on a surface visibility algorithm [153], typically used for viewpoint preference selection [154].

Overall, image-based metrics are not practical for driving processing operations (e.g., mesh simplification). Model-based metrics are better suited instead, since they operate on the same representation space with the corresponding processing algorithms; thus it is possible to control processing operations both globally (on the entire model) and locally (on the vertex/point level). At the same time, they typically require complex processes to effectively capture perceptually relevant features. Moreover, it is not straightforward

how to fuse information extracted from different attributes (e.g., geometry, texture) to obtain a total quality score.

Last but not least, the performance of image-based metrics greatly depends on distortions and contents. They are less accurate in differentiating and ranking different distortions, or distortions applied to different 3D models, which is not the case for model-based metrics [39,103].

18.2.4 Objective quality assessment in volumetric video

As video is a collection of still frames, volumetric video is also a collection of 3D models aggregated together, which are played back at a certain frame rate to create the motion perception. Temporal sub-sampling rate (i.e., frame rate) can be defined as the frequency of the consecutive models in the temporal axis of the volumetric video sequence. Utilizing all available frames in the volumetric video sequence is the common approach for objective quality evaluation. After predicting quality of each frame in the sequence, a temporal pooling method (e.g., arithmetic mean) is necessary to merge individual frame scores into a final quality score [23,24]. This is also commonly done in traditional video quality assessment, while extending the quality metrics that are developed for images to video [155]. However, one of the challenges for objective quality assessment of volumetric video is that the sizes of volumetric video sequences are big, and estimating objective quality can be time-consuming and computationally heavy. Reducing temporal sub-sampling rate and choosing appropriate pooling strategy may reduce the computational complexity without sacrificing from the prediction accuracy.

In a recent study, Ak et al. [156] investigated the performance of 30 quality metrics for 7 different temporal sampling methodologies over 8 different temporal sub-sampling rates. The study was conducted on the VsenseVVDB2 dataset [8], only on the point cloud sequences. The VsenseVVDB2 dataset contains 8 point cloud volumetric video sequences of 10 seconds length with 30 frames per second (fps). The utilized pooling methods are summarized in Table 18.4. Each pooling method was used with the following 8 different frame rates: {1, 2, 3, 5, 6, 10, 15, 30}. The frame rates were selected to ensure a uniform sampling.

Results are presented in Fig. 18.8, where SROCC is used to measure the performance of objective quality metrics. 11 image-based, 19 model-based metrics were evaluated on 56 combinations of 7 sub-sampling methods and 8 sub-sampling frequencies. Lighter colors indicate higher correlation with subjective opinions. Each row corresponds to a different quality metric indicated by the row number. Each column shows a different combination of temporal sub-sampling frequency and temporal sub-sampling method. Columns are divided into 7 groups by the sub-sampling method indicated at the bottom of the figure. For each sub-sampling method, frame rate increases from left to right.

Table 18.4 Definitions and selected parameters for pooling methods.

Sampling method	Formula	Parameter
Arithmetic mean	$Q = \frac{1}{N}\sum_{i=1}^{N} q_i$	-
Harmonic mean	$Q = \left(\frac{1}{N}\sum_{i=1}^{N} q_i^{-1}\right)^{-1}$	-
Minkowski mean	$Q = \left(\frac{1}{N}\sum_{i=1}^{N} q_i^{p}\right)^{1/p}$	$p = 2$
VQ pooling	$Q = \frac{\sum_{i\in G_L} q_i + w \cdot \sum_{i\in G_H} q_i}{\lvert G_L\rvert + w \cdot \lvert G_H\rvert}$, $w = \left(1 - \frac{M_L}{M_H}\right)^2$	-
Percentile pooling	$Q = \frac{1}{\lvert P_{low}\rvert}\sum_{i\in P_{low}} q_i$	Percentile = 10%
Primacy pooling	$Q = \sum_{i=1}^{N} w_i q_i$, $w_i = \frac{\exp(-\alpha i)}{\sum_{j=1}^{L}\exp(-\alpha j)}$, $0 \le i \le L$	$L = 360$, $\alpha = 0.01$
Recency pooling	$Q = \sum_{i=1}^{N} w_i q_i$, $w_i = \frac{\exp(-\alpha(L-i))}{\sum_{j=1}^{L}\exp(-\alpha(L-j))}$, $0 \le i \le L$	$L = 360$, $\alpha = 0.01$

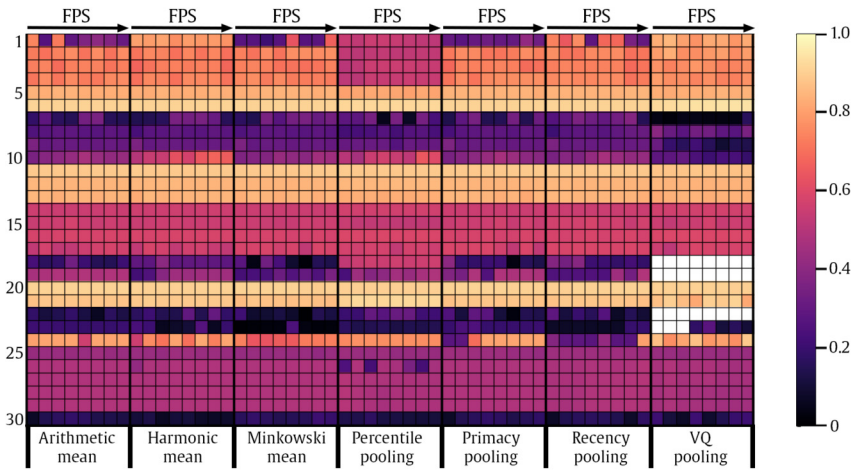

Figure 18.8 Metric performances in terms of SROCC for a number of sampling methodologies and sub-sampling rates. 1) MP-PSNR-FR, 2) MP-PSNR-RR, 3) MW-PSNR-FR, 4) MW-PSNR-RR, 5) PSNR, 6) SSIM, 7) NIQSV, 8) NIQSV+, 9) APT, 10) EM-IQM, 11) SI-IQM, 12) Color-Y, 13) Color-Y-PSNR, 14) Color-U, 15) Color-U-PSNR, 16) Color-V, 17) Color-V-PSNR, 18) point-to-point-Haus, 19) point-to-point-Haus-PSNR, 20) point-to-point-RMS, 21) point-to-point-RMS-PSNR, 22) point-to-plane-Haus, 23) point-to-plane-Haus-PSNR, 24) point-to-plane-RMS, 25) point-to-plane-RMS-PSNR, 26) plane-to-plane-MSE, 27) plane-to-plane-RMS, 28) plane-to-plane-Mean, 29) plane-to-plane-Median, 30) plane-to-plane-Min.

Results indicate, for any given temporal sampling method, using a lower frame-rate does not result in a lower performance. Even by sub-sampling by 1 fps, metrics performances do not show a significant difference compared to the full frame rate available (i.e., 30 fps). This observation indicates that compression artifacts affect the perceived quality of the volumetric video uniformly in time. For similar distortions, with no significant loss in the accuracy of objective quality metrics, calculations can be sped up to 30 times for stimuli.

Each considered pooling method has a different priority for the temporal dimension. Results in Fig. 18.8 show non-significant changes in metric performances with various pooling methods. Similar to the sub-sampling rate analysis, this occurs due to uniform presence of the compression distortions on the point clouds in the volumetric video sequences. Considering the non-significant performance difference of the quality metrics for various temporal sampling methods, arithmetic mean is the most efficient alternative due to computational simplicity.

It is worth noting that only point clouds of human bodies are used in this study, and early trials show that these findings might be valid for meshes and other types of 3D graphics as well. Although this needs further experimentation and validation by peer-review processes, using a reduced number of 3D models without sacrificing the metric accuracy can be very beneficial for wide deployment of volumetric video streaming.

18.2.5 Publicly available software implementations

As described in previous subsections, there are many different approaches in estimating the visual quality of volumetric video. In most cases, during the development of new objective quality metrics, a comparison is needed to validate the newly developed metrics's performance. To help future scientists and developers in their goals to propose more accurate quality metrics, publicly available implementations of existing methods are listed in Table 18.5. As can be seen, objective quality metrics for volumetric video can be grouped in different categories, with respect to 3D representation (i.e., mesh or point cloud); whether they demand color attributes or not; metric class in terms of reference data requirement; and domain of operation (i.e., model or image). Furthermore, a brief description of the features that each metric relies upon are indicated, along with a link to the open-source code.

18.3. Conclusion

Volumetric video is a novel form of visual representation that enables us to view at reconstructed 3D models from any viewpoint, which brings different challenges and limitations for visual quality assessment. The most important challenges are in understanding and estimating user interaction, selecting the correct 3D representation, and setting various conditions in applications such as rendering parameters and display

Table 18.5 Publicly available objective quality metric implementations for meshes & point clouds.

Metric	3D Representation	Attributes	Class	Domain	Features	Open-source code
Metro (1998) [102]	Mesh	Colorless	FR	Model-based	Mean error	https://sourceforge.net/projects/vcg/
Mesh (2002) [101]	Mesh	Colorless	FR	Model-based	Hausdorff distance	https://github.com/arnaudgelas/mesh
MSDM2 (2011) [54,66]	Mesh	Colorless	FR	Model-based	Local differences in curvature statistics	https://github.com/MEPP-team/MEPP2
TPDM (2014) [107]	Mesh	Colorless	FR	Model-based	Local differences in curvature tensor	http://www.gipsa-lab.fr/~fakhri.torkhani/software/TPDM.rar
DAME (2012) [35]	Mesh	Colorless	FR	Model-based	local differences in dihedral angles	http://meshcompression.org/software-tools
FMPD (2012) [108]	Mesh	Colorless	FR	Model-based	Global roughness using Gaussian curvature	http://www.gipsa-lab.grenoble-inp.fr/~kai.wang/publications_en.html
JND (2016) [109]	Mesh	Colorless	FR	Model-based	(bottom-up) visibility threshold predictor based on local contrast and spatial frequency	https://github.com/MEPP-team/MEPP2
Yildiz et al. (2020) [113]	Mesh	Colorless	FR	Model-based	Learning geometric parameters of a distance metric using crowd-sourced data	https://www.dropbox.com/s/m3bnb93vun91763/Learning.VQA.zip?dl=0

continued on next page

Table 18.5 (*continued*)

Metric	3D Representation	Attributes	Class	Domain	Features	Open-source code
STED (2011) [42]	Dynamic mesh	Colorless	FR	Model-based	Local differences in edge length (spatial and temporal parts)	http://meshcompression.org/software-tools
CMDM (2021) [39]	Mesh	Colored	FR	Model-based	Local differences in curvature and color statistics	https://github.com/MEPP-team/MEPP2
Graphics-LPIPS (2022) [41]	Mesh	Colored	FR	Image-based	CNN with linear weights on top	https://github.com/YanaNEHME/Graphics-LPIPS
Plane-to-plane (2018) [73]	Point cloud	Colorless	FR	Model-based	Angular similarity between normal vectors	https://github.com/mmspg/point-cloud-angular-similarity-metric
PC-MSDM (2019) [77]	Point cloud	Colorless	FR	Model-based	Curvature statistics	https://github.com/MEPP-team/PC-MSDM
PCQM (2020) [78]	Point cloud	Colored	FR	Model-based	Curvature and color statistics	https://github.com/MEPP-team/PCQM
Hist_Y (2020) [92]	Point cloud	Colored	FR	Model-based	Luminance histogram	https://github.com/cwi-dis/point-cloud-color-metric
PointSSIM (2020) [79]	Point cloud	Colored	FR	Model-based	Location, angular similarity, curvature, or luminance statistics	https://github.com/mmspg/pointssim
Point-to-distribution (2020) [75]	Point cloud	Colored	FR	Model-based	Mahalanobis distance of point coordinates and luminance	https://github.com/AlirezaJav/Point_to_distribution_metric

continued on next page

Table 18.5 (*continued*)

Metric	3D Representation	Attributes	Class	Domain	Features	Open-source code
PCM_RR (2020) [93]	Point cloud	Colored	RR	Model-based	Location, angular similarity and luminance histograms	https://github.com/cwi-dis/PCM_RR
GraphSIM (2020) [83]	Point cloud	Colored	FR	Model-based	Color gradient statistics around keypoints using graphs	https://github.com/NJUVISION/GraphSIM
Perceptual loss (2021) [98]	Point cloud	Colorless	FR	Model-based	Differences in latent space after auto-encoding voxel grids with binary or truncated distance fields	https://github.com/mauriceqch/2021_pc_perceptual_loss
BitDance (2021) [91]	Point cloud	Colored	FR	Model-based	Bit-based differences of colors using CIEDE2000 and normal vectors	https://github.com/rafael2k/bitdance-pc_metric
MS-GraphSIM (2021) [84]	Point cloud	Colored	FR	Model-based	Color gradient statistics on multi-scale representations around keypoints using graphs	https://github.com/zyj1318053/MS_GraphSIM
PointPCA (2021) [82]	Point cloud	Colored	FR	Model-based	Statistics on PCA-based geometric descriptors and luminance	https://github.com/cwi-dis/pointpca
PQA-Net (2021) [145]	Point cloud	Colored	NR	Image-based	Multi-view CNN-based features fed to a distortion identifier and a quality predictor	https://github.com/qdushl/PQA-Net

devices. This chapter provides a wide overview of quality assessment and estimation methodologies through subjective user studies and objective quality estimators for volumetric content.

With the increased degrees of freedom, users need to interact with the media itself. This makes user interaction a crucial part of subjective quality assessment. Nevertheless, there are no recommendations or standards for conducting user studies for volumetric video in 6DoF. Therefore experimenters adapted various methodologies for subjective user studies that were developed for traditional image and video quality assessment with different parameters for various aspects. These involve the 3D representation, temporal variation, attributes, mode of inspection, subjective test methodology, distortion types, rendering parameters, and display devices, among others. Ground truth subjective quality scores were collected as part of these studies, with publicly released datasets reported in this chapter. Comparative studies show that there are certain cases that selecting one approach is more efficient than others. When compressed with state-of-the-art codecs, point clouds seem to be better at low-bandwidth conditions, whereas meshes are better at high-bandwidth conditions or storage cases. Interactive and non-interactive experiments generally do not have statistically significant differences. Nevertheless, interactive visualization approaches might be more suitable, as they better simulate targeted use cases, with human judgments implicitly incorporating effects of display devices, rendering parameters, higher DoF, and immersion, to name a few. Among the commonly used subjective methodologies, the DSIS seems to be more accurate by yielding lower uncertainty regarding the level of impairment perceived in a stimulus. However, such a methodology may lead to comparative scores.

Automatic estimation of volumetric video quality via objective quality metrics is still under development. Model-based approaches rely on the primary 3D data structure. The unique nature of volumetric video enables its representation with meshes and point clouds. Since point clouds do not have connectivity information, the majority of corresponding metrics focus on capturing underlying 3D surfaces. Early attempts for both point clouds and meshes are based on simple error measurements, whereas more recent efforts combine various features from geometry and/or texture domain. Image-based approaches, on the other hand, focus on projecting the 3D model onto planar arrangements and often making use of 2D quality metrics. Comparative studies show both approaches have different advantages and disadvantages. Although image-based approaches require rendering and camera parameters to be set beforehand, they take all the rendering effects into account, while estimating the quality. Model-based approaches generate scores that are independent of the viewpoint; however, they need to be rather complex to take human visual perception into account, which usually results in high computational demands. The selection of either approach depends on the application, as they both have strengths and weaknesses. Recent studies also show that to estimate volumetric video quality for compression scenarios, we do not need to compute the metric results for all of the frames (i.e., consecutive 3D models in a volumetric

video). Since most compression methods generate distortions that do not change by time, selecting fewer frames yields as accurate quality estimations as computing it over all frames, although further experimentation is required at this front.

With the popularization of XR technologies and applications on the Internet, social media, and metaverse(s), the volumetric video will become more popular. Volumetric video will be used in different mixed XR applications alongside other types of 3D graphics, and there will be a need for new methodologies to capture the human perception and predict human opinions about visual quality within these mixed-media environments. This makes the field of visual quality assessment very relevant, and open to the new challenges the upcoming advances will bring.

References

[1] A. Perkis, C. Timmerer, S. Baraković, J. Baraković Husić, S. Bech, S. Bosse, J. Botev, K. Brunnström, L. Cruz, K. De Moor, A. de Polo Saibanti, W. Durnez, S. Egger-Lampl, U. Engelke, T.H. Falk, J. Gutiérrez, A. Hameed, A. Hines, T. Kojic, D. Kukolj, E. Liotou, D. Milovanovic, S. Möller, N. Murray, B. Naderi, M. Pereira, S. Perry, A. Pinheiro, A. Pinilla, A. Raake, S.R. Agrawal, U. Reiter, R. Rodrigues, R. Schatz, P. Schelkens, S. Schmidt, S.S. Sabet, A. Singla, L. Skorin-Kapov, M. Suznjevic, S. Uhrig, S. Vlahović, J.-N. Voigt-Antons, S. Zadtootaghaj, QUALINET white paper, on definitions of immersive media experience (IMEx), European Network on Quality of Experience in Multimedia, Systems and Services, 14th QUALINET meeting (online), https://arxiv.org/abs/2007.07032, May 2020.

[2] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P.A. Chou, R.A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A.M. Tourapis, V. Zakharchenko, Emerging MPEG standards for point cloud compression, IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9 (1) (2019) 133–148, https://doi.org/10.1109/JETCAS.2018.2885981.

[3] T. Ebrahimi, S. Foessel, F. Pereira, P. Schelkens, JPEG Pleno: Toward an efficient representation of visual reality, IEEE MultiMedia 23 (4) (2016) 14–20, https://doi.org/10.1109/MMUL.2016.64.

[4] ITU-T, Subjective video quality assessment methods for multimedia applications, ITU-T Recommendation P.910, Apr 2008.

[5] ITU-R, Methodology for the subjective assessment of the quality of television pictures, ITU-R Recommendation BT.500, Jan 2012.

[6] F. Kozamernik, V. Steinmann, P. Sunna, E. Wyckens, SAMVIQ—A new EBU methodology for video quality evaluations in multimedia, SMPTE Motion Imaging Journal 114 (4) (2005) 152–160.

[7] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al., Image database TID2013: Peculiarities, results and perspectives, Signal Processing. Image Communication 30 (2015) 57–77.

[8] E. Zerman, C. Ozcinar, P. Gao, A. Smolic, Textured mesh vs coloured point cloud: A subjective study for volumetric video compression, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2020, pp. 1–6.

[9] P. Le Callet, S. Möller, A. Perkis, Qualinet white paper on definitions of quality of experience, European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.2, Mar 2013.

[10] ITU-T, Subjective test methodologies for 360° video on head-mounted displays, ITU-T Recommendation P.919, Oct 2020.

[11] J. Gutierrez, P. Perez, M. Orduna, A. Singla, C. Cortes, P. Mazumdar, I. Viola, K. Brunnstrom, F. Battisti, N. Cieplinska, et al., Subjective evaluation of visual quality and simulator sickness of short 360 videos: ITU-T Rec, IEEE Transactions on Multimedia (2021) 919.

[12] ITU-T, Subjective test method for interactive virtual reality applications, ITU-T Work Programme P.IntVr, https://www.itu.int/itu-t/workprog/wp_item.aspx?isn=17045. (Accessed 19 January 2022).
[13] ITU-T, QoE assessment of extended reality (XR) meetings, ITU-T Work Programme P.QXM, https://www.itu.int/itu-t/workprog/wp_item.aspx?isn=15113. (Accessed 19 January 2022).
[14] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, Subjective and objective quality evaluation of compressed point clouds, in: 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), 2017, pp. 1–6, https://doi.org/10.1109/MMSP.2017.8122239.
[15] H. Su, Z. Duanmu, W. Liu, Q. Liu, Z. Wang, Perceptual quality assessment of 3D point clouds, in: 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 3182–3186.
[16] D. Lazzarotto, E. Alexiou, T. Ebrahimi, Benchmarking of objective quality metrics for point cloud compression, in: 2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP), 2021, pp. 1–6.
[17] L.A. da Silva Cruz, E. Dumić, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, T. Ebrahimi, Point cloud quality evaluation: Towards a definition for test conditions, in: 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), 2019, pp. 1–6, https://doi.org/10.1109/QoMEX.2019.8743258.
[18] S. Perry, H.P. Cong, L.A. da Silva Cruz, J. Prazeres, M. Pereira, A. Pinheiro, E. Dumic, E. Alexiou, T. Ebrahimi, Quality evaluation of static point clouds encoded using MPEG codecs, in: 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 3428–3432, https://doi.org/10.1109/ICIP40778.2020.9191308.
[19] E. Alexiou, T. Ebrahimi, M.V. Bernardo, M. Pereira, A. Pinheiro, L.A. Da Silva Cruz, C. Duarte, L.G. Dmitrovic, E. Dumic, D. Matkovics, A. Skodras, Point cloud subjective evaluation methodology based on 2D rendering, in: 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), 2018, pp. 1–6, https://doi.org/10.1109/QoMEX.2018.8463406.
[20] E. Alexiou, A.M.G. Pinheiro, C. Duarte, D. Matković, E. Dumić, L.A. da Silva Cruz, L.G. Dmitrović, M.V. Bernardo, M. Pereira, T. Ebrahimi, Point cloud subjective evaluation methodology based on reconstructed surfaces, in: A.G. Tescher (Ed.), Applications of Digital Image Processing XLI, vol. 10752, International Society for Optics and Photonics, SPIE, 2018, pp. 160–173.
[21] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, Point cloud rendering after coding: impacts on subjective and objective quality, arXiv preprint, arXiv:1912.09137, 2019.
[22] R. Mekuria, S. Laserre, C. Tulvan, Performance assessment of point cloud compression, in: 2017 IEEE Visual Communications and Image Processing (VCIP), 2017, pp. 1–4, https://doi.org/10.1109/VCIP.2017.8305132.
[23] Emin Zerman, Pan Gao, Cagri Ozcinar, Aljosa Smolic, Subjective and objective quality assessment for volumetric video compression, in: Proc. IS&T Int'l. Symp. on Electronic Imaging: Image Quality and System Performance XVI, 2019, pp. 323-1–323-7, https://doi.org/10.2352/ISSN.2470-1173.2019.10.IQSP-323.
[24] J. van der Hooft, M.T. Vega, C. Timmerer, A.C. Begen, F. De Turck, R. Schatz, Objective and subjective QoE evaluation for adaptive point cloud streaming, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6.
[25] E. Alexiou, T. Ebrahimi, On the performance of metrics to predict quality in point cloud representations, in: A.G. Tescher (Ed.), Applications of Digital Image Processing XL, vol. 10396, International Society for Optics and Photonics, SPIE, 2017, pp. 282–297.
[26] E. Alexiou, E. Upenik, T. Ebrahimi, Towards subjective quality assessment of point cloud imaging in augmented reality, in: 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), 2017, pp. 1–6, https://doi.org/10.1109/MMSP.2017.8122237.
[27] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, Subjective and objective quality evaluation of 3D point cloud denoising algorithms, in: 2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2017, pp. 1–6.
[28] E. Alexiou, T. Ebrahimi, Exploiting user interactivity in quality assessment of point cloud imaging, in: 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), 2019, pp. 1–6, https://doi.org/10.1109/QoMEX.2019.8743277.

[29] E. Alexiou, I. Viola, T.M. Borges, T.A. Fonseca, R.L. de Queiroz, T. Ebrahimi, A comprehensive study of the rate-distortion performance in MPEG point cloud compression, APSIPA Transactions on Signal and Information Processing 8 (2019) e27.
[30] M. Kazhdan, H. Hoppe, Screened Poisson surface reconstruction, ACM Transactions on Graphics 32 (3) (2013) 29, 13pp.
[31] MPEG 3DG and Requirements, Call for proposals for point cloud compression v2, ISO/IEC JTC1/SC29/WG11 Doc. N16763, Apr. 2017.
[32] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, H. Hellwagner, Towards 6DoF HTTP adaptive streaming through point cloud compression, in: Proceedings of the 27th ACM International Conference on Multimedia, MM '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 2405–2413.
[33] Benjamin Watson, Alinda Friedman, Aaron McGaffey, Measuring and predicting visual fidelity, in: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '01, Association for Computing Machinery, New York, NY, USA, 2001, pp. 213–220, https://doi.org/10.1145/383259.383283.
[34] B.E. Rogowitz, H.E. Rushmeier, Are image quality metrics adequate to evaluate the quality of geometric objects?, in: Human Vision and Electronic Imaging VI, in: Proc. SPIE, vol. 4299, 2001, https://doi.org/10.1117/12.429504.
[35] L. Váša, J. Rus, Dihedral Angle Mesh Error: a fast perception correlated distortion measure for fixed connectivity triangle meshes, Computer Graphics Forum 31 (5) (2012).
[36] Y. Pan, I. Cheng, A. Basu, Quality metric for approximating subjective evaluation of 3-D objects, IEEE Transactions on Multimedia 7 (2) (2005) 269–279, https://doi.org/10.1109/TMM.2005.843364.
[37] J. Guo, V. Vidal, I. Cheng, A. Basu, A. Baskurt, G. Lavoué, Subjective and objective visual quality assessment of textured 3D meshes, ACM Transactions on Applied Perception 14 (2016) 1–20, https://doi.org/10.1145/2996296.
[38] A. Doumanoglou, N. Zioulis, E. Christakis, D. Zarpalas, P. Daras, Subjective quality assessment of textured human full-body 3D-reconstructions, in: 2018 10th International Conference on Quality of Multimedia Experience, QoMEX 2018, 2018.
[39] Y. Nehmé, F. Dupont, J.-P. Farrugia, P. Le Callet, G. Lavoué, Visual quality of 3D meshes with diffuse colors in virtual reality: Subjective and objective evaluation, IEEE Transactions on Visualization and Computer Graphics 27 (3) (2021) 2202–2219, https://doi.org/10.1109/TVCG.2020.3036153.
[40] Y. Nehmé, J.-P. Farrugia, F. Dupont, P.L. Callet, G. Lavoué, Comparison of subjective methods for quality assessment of 3D graphics in virtual reality, ACM Transactions on Applied Perception 18 (1) (2020) 1–23, https://doi.org/10.1145/3427931.
[41] Y. Nehmé, F. Dupont, J.-P. Farrugia, P. Le Callet, G. Lavoué, Textured mesh quality assessment: Large-scale dataset and deep learning-based quality metric, arXiv preprint, arXiv:2202.02397, 2022.
[42] L. Váša, V. Skala, A perception correlated comparison method for dynamic meshes, IEEE Transactions on Visualization and Computer Graphics 17 (2011) 220–230, https://doi.org/10.1109/TVCG.2010.38.
[43] F. Torkhani, K. Wang, J.-M. Chassery, Perceptual quality assessment of 3D dynamic meshes: Subjective and objective studies, Signal Processing. Image Communication 31 (2) (2015) 185–204, https://doi.org/10.1016/j.image.2014.12.008.
[44] K. Vanhoey, B. Sauvage, P. Kraemer, G. Lavoué, Visual quality assessment of 3D models: On the influence of light-material interaction, ACM Transactions on Applied Perception 15 (1) (2017), https://doi.org/10.1145/3129505.
[45] Y. Nehmé, P.L. Callet, F. Dupont, J.-P. Farrugia, G. Lavoué, Exploring crowdsourcing for subjective quality assessment of 3D graphics, in: IEEE International Workshop on Multimedia Signal Processing (MMSP), 2021.
[46] E. Alexiou, T. Ebrahimi, On subjective and objective quality evaluation of point cloud geometry, in: 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), 2017, pp. 1–3, https://doi.org/10.1109/QoMEX.2017.7965681.
[47] E.M. Torlig, E. Alexiou, T.A. Fonseca, R.L. de Queiroz, T. Ebrahimi, A novel methodology for quality assessment of voxelized point clouds, in: A.G. Tescher (Ed.), Applications of Digital

Image Processing XLI, vol. 10752, International Society for Optics and Photonics, SPIE, 2018, pp. 174–190.
[48] Qi Yang, Hao Chen, Zhan Ma, Yiling Xu, Rongjun Tang, Jun Sun, Predicting the perceptual quality of point cloud: A 3D-to-2D projection-based exploration, IEEE Transactions on Multimedia 23 (2020) 3877–3891, https://doi.org/10.1109/TMM.2020.3033117.
[49] R. Mekuria, K. Blom, P. Cesar, Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video, IEEE Transactions on Circuits and Systems for Video Technology 27 (4) (2017) 828–842.
[50] E. Alexiou, N. Yang, T. Ebrahimi, PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6, https://doi.org/10.1109/QoMEX48832.2020.9123121.
[51] S. Subramanyam, J. Li, I. Viola, P. Cesar, Comparing the quality of highly realistic digital humans in 3DoF and 6DoF: A volumetric video case study, in: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE, 2020, pp. 127–136.
[52] X. Wu, Y. Zhang, C. Fan, J. Hou, S. Kwong, Subjective quality database and objective study of compressed point clouds with 6DoF head-mounted display, IEEE Transactions on Circuits and Systems for Video Technology 31 (12) (2021) 4630–4644.
[53] I. Viola, S. Subramanyam, J. Li, P. Cesar, On the impact of VR assessment on the quality of experience of highly realistic digital humans, Quality and User Experience 7 (1) (2022) 3, https://link.springer.com/article/10.1007/s41233-022-00050-3.
[54] G. Lavoué, E. Drelie Gelasca, F. Dupont, A. Baskurt, T. Ebrahimi, Perceptually driven 3D distance metrics with application to watermarking, in: Applications of Digital Image Processing XXIX, vol. 6312, 2006, pp. 150–161, https://doi.org/10.1117/12.686964.
[55] M. Corsini, E.D. Gelasca, T. Ebrahimi, M. Barni, Watermarked 3-D mesh quality assessment, IEEE Transactions on Multimedia 9 (2007) 247–256.
[56] G. Lavoué, A local roughness measure for 3D meshes and its application to visual masking, ACM Transactions on Applied Perception 5 (4) (2009), https://doi.org/10.1145/1462048.1462052.
[57] S. Silva, B.S. Santos, C. Ferreira, J. Madeira, A Perceptual Data Repository for Polygonal Meshes, in: 2009 Second International Conference in Visualisation, 2009, pp. 207–212.
[58] R. Mekuria, P. Cesar, I. Doumanis, A. Frisiello, Objective and subjective quality assessment of geometry compression of reconstructed 3D humans in a 3D virtual room, in: A.G. Tescher (Ed.), Applications of Digital Image Processing XXXVIII, vol. 9599, International Society for Optics and Photonics, SPIE, 2015, pp. 537–549, https://doi.org/10.1117/12.2203312.
[59] K. Christaki, E. Christakis, 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.
[60] J. Gutiérrez, T. Vigier, P. Le Callet, Quality evaluation of 3D objects in mixed reality for different lighting conditions, Electronic Imaging 2020 (2020), https://doi.org/10.2352/ISSN.2470-1173.2020.11.HVEI-128.
[61] T. Hoßfeld, P.E. Heegaard, M. Varela, S. Möller, QoE beyond the MOS: an in-depth look at QoE via better metrics and their relation to MOS, Quality and User Experience 1 (1) (2016) 1–23.
[62] E. Zerman, G. Valenzise, F. Dufaux, An extensive performance evaluation of full-reference HDR image quality metrics, Quality and User Experience 2 (1) (2017) 5.
[63] L. Krasula, K. Fliegel, P. Le Callet, M. Klíma, On the accuracy of objective image and video quality models: New methodology for performance evaluation, in: 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2016, pp. 1–6.
[64] K. Cao, Y. Xu, P. Cosman, Visual Quality of Compressed Mesh and Point Cloud Sequences, IEEE Access 8 (2020) 171203–171217.
[65] E. Alexiou, T. Ebrahimi, Impact of visualisation strategy for subjective quality assessment of point clouds, in: 2018 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2018, pp. 1–6, https://doi.org/10.1109/ICMEW.2018.8551498.
[66] G. Lavoué, A multiscale metric for 3D mesh visual quality assessment, Computer Graphics Forum 30 (5) (2011) 1427–1437.

[67] D. Tian, H. Ochimizu, C. Feng, R. Cohen, A. Vetro, Geometric distortion metrics for point cloud compression, in: 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 3460–3464, https://doi.org/10.1109/ICIP.2017.8296925.
[68] MPEG 3DG Group, Common test conditions for point cloud compression, ISO/IEC JTC1/SC29/WG11 Doc. N18474, Mar. 2019.
[69] D. Tian, H. Ochimizu, C. Feng, R. Cohen, A. Vetro, Updates and integration of evaluation metric software for PCC, ISO/IEC JTC1/SC29/WG11 Doc. MPEG2017/M40522, Apr. 2017.
[70] D. Tian, H. Ochimizu, C. Feng, R. Cohen, A. Vetro, Evaluation metrics for point cloud compression, ISO/IEC JTC1/SC29/WG11 Doc. M39966, Jan. 2017.
[71] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, A generalized Hausdorff distance based quality metric for point cloud geometry, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6, https://doi.org/10.1109/QoMEX48832.2020.9123087.
[72] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, Improving PSNR-based quality metrics performance for point cloud geometry, in: 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 3438–3442, https://doi.org/10.1109/ICIP40778.2020.9191233.
[73] E. Alexiou, T. Ebrahimi, Point cloud quality assessment metric based on angular similarity, in: 2018 IEEE International Conference on Multimedia and Expo (ICME), 2018, pp. 1–6, https://doi.org/10.1109/ICME.2018.8486512.
[74] E. Alexiou, T. Ebrahimi, Benchmarking of the plane-to-plane metric, ISO/IEC JTC1/SC29/WG1 Doc. M88038, Jul. 2020.
[75] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, Mahalanobis based point to distribution metric for point cloud geometry quality evaluation, IEEE Signal Processing Letters 27 (2020) 1350–1354, https://doi.org/10.1109/LSP.2020.3010128.
[76] A. Javaheri, C. Brites, F. Pereira, J. Ascenso, A point-to-distribution joint geometry and color metric for point cloud quality assessment, arXiv preprint, arXiv:2108.00054, 2021.
[77] G. Meynet, J. Digne, G. Lavoué, PC-MSDM: A quality metric for 3D point clouds, in: 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), 2019, pp. 1–3, https://doi.org/10.1109/QoMEX.2019.8743313.
[78] G. Meynet, Y. Nehmé, J. Digne, G. Lavoué, PCQM: A full-reference quality metric for colored 3D point clouds, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6, https://doi.org/10.1109/QoMEX48832.2020.9123147.
[79] E. Alexiou, T. Ebrahimi, Towards a point cloud structural similarity metric, in: 2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2020, pp. 1–6, https://doi.org/10.1109/ICMEW46912.2020.9106005.
[80] L. Hua, M. Yu, G. Jiang, Z. He, Y. Lin, VQA-CPC: a novel visual quality assessment metric of color point clouds, in: Q. Dai, T. Shimura, Z. Zheng (Eds.), Optoelectronic Imaging and Multimedia Technology VII, vol. 11550, International Society for Optics and Photonics, SPIE, 2020, pp. 244–252.
[81] Lei Hua, Mei Yu, Zhouyan He, Renwei Tu, Gangyi Jiang, CPC-GSCT: Visual quality assessment for coloured point cloud based on geometric segmentation and colour transformation, IET Image Processing 16 (4) (2022) 1083–1095, https://doi.org/10.1049/ipr2.12211.
[82] E. Alexiou, I. Viola, P. Cesar, PointPCA: Point cloud objective quality assessment using PCA-based descriptors, arXiv preprint, arXiv:2111.12663, 2021.
[83] Qi Yang, Zhan Ma, Yiling Xu, Zhu Li, Jun Sun, Inferring point cloud quality via graph similarity, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (6) (2022) 3015–3029, https://doi.org/10.1109/TPAMI.2020.3047083.
[84] Y. Zhang, Q. Yang, Y. Xu, MS-GraphSIM: Inferring Point Cloud Quality via Multiscale Graph Similarity, Association for Computing Machinery, New York, NY, USA, 2021, pp. 1230–1238, https://doi.org/10.1145/3474085.3475294.
[85] K.-x. Zhang, G.-y. Jiang, M. Yu, FQM-GC: Full-reference quality metric for colored point cloud based on graph signal features and color features, in: ACM Multimedia Asia, MMAsia '21, Association for Computing Machinery, New York, NY, USA, 2021, https://doi.org/10.1145/3469877.3490578.

[86] Y. Xu, Q. Yang, L. Yang, J.-N. Hwang, EPES: Point cloud quality modeling using elastic potential energy similarity, IEEE Transactions on Broadcasting (2021) 1–10, https://doi.org/10.1109/TBC.2021.3114510.
[87] R. Diniz, P.G. Freitas, M.C.Q. Farias, Towards a point cloud quality assessment model using local binary patterns, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6, https://doi.org/10.1109/QoMEX48832.2020.9123076.
[88] R. Diniz, P.G. Freitas, M.C. Farias, Multi-distance point cloud quality assessment, in: 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 3443–3447, https://doi.org/10.1109/ICIP40778.2020.9190956.
[89] R. Diniz, P.G. Freitas, M.C. Farias, Local luminance patterns for point cloud quality assessment, in: 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP), 2020, pp. 1–6, https://doi.org/10.1109/MMSP48831.2020.9287154.
[90] R. Diniz, P.G. Freitas, M. Farias, A novel point cloud quality assessment metric based on perceptual color distance patterns, Electronic Imaging 2021 (9) (2021) 256, 11pp., https://doi.org/10.2352/ISSN.2470-1173.2021.9.IQSP-256.
[91] R. Diniz, P.G. Freitas, M.C.Q. Farias, Color and geometry texture descriptors for point-cloud quality assessment, IEEE Signal Processing Letters 28 (2021) 1150–1154, https://doi.org/10.1109/LSP.2021.3088059.
[92] I. Viola, S. Subramanyam, P. Cesar, A color-based objective quality metric for point cloud contents, in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6, https://doi.org/10.1109/QoMEX48832.2020.9123089.
[93] I. Viola, P. Cesar, A reduced reference metric for visual quality evaluation of point cloud contents, IEEE Signal Processing Letters 27 (2020) 1660–1664, https://doi.org/10.1109/LSP.2020.3024065.
[94] Q. Liu, H. Yuan, R. Hamzaoui, H. Su, J. Hou, H. Yang, Reduced reference perceptual quality model with application to rate control for video-based point cloud compression, IEEE Transactions on Image Processing 30 (2021) 6623–6636, https://doi.org/10.1109/TIP.2021.3096060.
[95] L. Hua, G. Jiang, M. Yu, Z. He, BQE-CVP: Blind quality evaluator for colored point cloud based on visual perception, in: 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), 2021, pp. 1–6, https://doi.org/10.1109/BMSB53066.2021.9547070.
[96] X. Yang, W. Ling, Z. Lu, E. Ong, S. Yao, Just noticeable distortion model and its applications in video coding, Signal Processing. Image Communication 20 (7) (2005) 662–680, https://doi.org/10.1016/j.image.2005.04.001.
[97] A. Chetouani, M. Quach, G. Valenzise, F. Dufaux, Deep learning-based quality assessment of 3D point clouds without reference, in: 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2021, pp. 1–6, https://doi.org/10.1109/ICMEW53276.2021.9455967.
[98] M. Quach, A. Chetouani, G. Valenzise, F. Dufaux, A deep perceptual metric for 3D point clouds, Electronic Imaging 2021 (9) (2021) 257, 7pp., https://doi.org/10.2352/ISSN.2470-1173.2021.9.IQSP-257.
[99] M. Corsini, M.C. Larabi, G. Lavoué, O. Petrik, L. Váša, K. Wang, Perceptual Metrics for Static and Dynamic Triangle Meshes, Computer Graphics Forum 32 (1) (2013) 101–125.
[100] G. Lavoué, R. Mantiuk, Quality assessment in computer graphics, in: Visual Signal Quality Assessment: Quality of Experience (QoE), 2015, pp. 243–286.
[101] N. Aspert, D. Santa-Cruz, T. Ebrahimi, MESH: measuring errors between surfaces using the Hausdorff distance, in: Proceedings. IEEE International Conference on Multimedia and Expo, 2002, https://doi.org/10.1109/ICME.2002.1035879.
[102] P. Cignoni, C. Rocchini, R. Scopigno, Metro: Measuring error on simplified surfaces, Computer Graphics Forum 17 (1998).
[103] G. Lavoué, M.C. Larabi, L. Vasa, On the Efficiency of Image Metrics for Evaluating the Visual Quality of 3D Models, IEEE Transactions on Visualization and Computer Graphics 22 (8) (2016) 1987–1999.
[104] Z. Karni, C. Gotsman, Spectral compression of mesh geometry, in: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, 2000, pp. 279–286, https://doi.org/10.1145/344779.344924.

[105] Z. Bian, S.-M. Hu, R. Martin, Evaluation for small visual difference between conforming meshes on strain field, Journal of Computer Science and Technology 24 (2009), https://doi.org/10.1007/s11390-009-9198-3.
[106] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.
[107] F. Torkhani, K. Wang, J.-M. Chassery, A curvature-tensor-based perceptual quality metric for 3D triangular meshes, Machine Graphics & Vision 23 (1) (2014) 59–82.
[108] Kai Wang, Fakhri Torkhani, Annick Montanvert, A fast roughness-based approach to the assessment of 3D mesh visual quality, Computers & Graphics 36 (7) (2012) 808–818, https://doi.org/10.1016/j.cag.2012.06.004.
[109] G. Nader, K. Wang, F. Hétroy-Wheeler, F. Dupont, Just noticeable distortion profile for flat-shaded 3D mesh surfaces, IEEE Transactions on Visualization and Computer Graphics 22 (11) (2016) 2423–2436, https://doi.org/10.1109/TVCG.2015.2507578.
[110] J. Guo, V. Vidal, A. Baskurt, G. Lavou, Evaluating the local visibility of geometric artifacts, in: ACM Symposium in Applied Perception, 2015.
[111] G. Lavoué, I. Cheng, A. Basu, Perceptual quality metrics for 3D meshes: Towards an optimal multi-attribute computational model, in: 2013 IEEE International Conference on Systems, Man, and Cybernetics, 2013, pp. 3271–3276, https://doi.org/10.1109/SMC.2013.557.
[112] A. Chetouani, Three-dimensional mesh quality metric with reference based on a support vector regression model, Journal of Electronic Imaging 27 (4) (2018) 1–9, https://doi.org/10.1117/1.JEI.27.4.043048.
[113] Z.C. Yildiz, A.C. Oztireli, T. Capin, A machine learning framework for full-reference 3D shape quality assessment, Visual Computer 36 (1) (2020) 127–139.
[114] Z.C. Yildiz, T. Capin, A perceptual quality metric for dynamic triangle meshes, EURASIP Journal on Image and Video Processing 2017 (12) (2017).
[115] I. Abouelaziz, M. El Hassouni, H. Cherifi, A curvature based method for blind mesh visual quality assessment using a general regression neural network, in: 2016 12th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), 2016, pp. 793–797, https://doi.org/10.1109/SITIS.2016.130.
[116] A. Nouri, C. Charrier, O. Lézoray, 3D Blind Mesh Quality Assessment Index, IS&T International Symposium on Electronic Imaging, Jan. 2017.
[117] I. Abouelaziz, M. El Hassouni, H. Cherifi, No-reference 3D mesh quality assessment based on dihedral angles model and support vector regression, Image and Signal Processing (2016) 369–377.
[118] I. Abouelaziz, M.E. Hassouni, H. Cherifi, A convolutional neural network framework for blind mesh visual quality assessment, in: 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 755–759, https://doi.org/10.1109/ICIP.2017.8296382.
[119] D. Tian, G. AlRegib, Batex3: Bit allocation for progressive transmission of textured 3-D models, IEEE Transactions on Circuits and Systems for Video Technology 18 (1) (2008) 23–35.
[120] Y. Zhao, Y. Liu, R. Song, M. Zhang, A saliency detection based method for 3D surface simplification, in: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 889–892, https://doi.org/10.1109/ICASSP.2012.6288027.
[121] X. Jiao, T. Wu, X. Qin, Mesh segmentation by combining mesh saliency with spectral clustering, Journal of Computational and Applied Mathematics 329 (2018) 134–146.
[122] S. Croci, S.B. Knorr, L. Goldmann, A. Smolic, A framework for quality control in cinematic VR based on Voronoi patches and saliency, in: 2017 International Conference on 3D Immersion (IC3D), 2017, pp. 1–8.
[123] S. Croci, C. Ozcinar, E. Zerman, S. Knorr, J. Cabrera, A. Smolic, Visual attention-aware quality estimation framework for omnidirectional video using spherical Voronoi diagram, Quality and User Experience 5 (2020), https://doi.org/10.1007/s41233-020-00032-3.
[124] Y. Nehmé, M. Abid, G. Lavoué, M.P.D. Silva, P.L. Callet, CMDM-VAC: Improving a perceptual quality metric for 3D graphics by integrating a visual attention complexity measure, in: 2021 IEEE International Conference on Image Processing (ICIP), 2021, pp. 3368–3372, https://doi.org/10.1109/ICIP42928.2021.9506662.

[125] M. Abid, M. Perreira Da Silva, P. Le Callet, Perceptual characterization of 3D graphical contents based on attention complexity measures, in: QoEVMA'20: Proceedings of the 1st Workshop on Quality of Experience (QoE) in Visual Multimedia Applications, 2020, pp. 31–36, https://doi.org/10.1145/3423328.3423498.

[126] J. Lubin, A visual discrimination model for imaging system design and evaluation, in: Vision Models for Target Detection and Recognition, 1995, pp. 245–283, https://doi.org/10.1142/9789812831200_0010.

[127] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Transactions on Image Processing 15 (2) (2006) 430–444.

[128] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Transactions on Image Processing 20 (8) (2011) 2378–2386, https://doi.org/10.1109/TIP.2011.2109730.

[129] R. Mantiuk, K.J. Kim, A.G. Rempel, W. Heidrich, HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions, ACM Transactions on Graphics 30 (4) (Jul. 2011), https://doi.org/10.1145/2010324.1964935.

[130] J. Preiss, F. Fernandes, P. Urban, Color-image quality assessment: From prediction to optimization, IEEE Transactions on Image Processing 23 (3) (2014) 1366–1378, https://doi.org/10.1109/TIP.2014.2302684.

[131] M.A. Saad, A.C. Bovik, C. Charrier, A DCT statistics-based blind image quality index, IEEE Signal Processing Letters 17 (6) (2010) 583–586, https://doi.org/10.1109/LSP.2010.2045550.

[132] W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE Transactions on Image Processing 23 (2) (2014) 684–695, https://doi.org/10.1109/TIP.2013.2293423.

[133] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, Y. Zhu, DeepSim: Deep similarity for image quality assessment, Neurocomputing 257 (2017) 104–114, https://doi.org/10.1016/j.neucom.2017.01.054, Machine Learning and Signal Processing for Big Multimedia Analysis.

[134] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 586–595, https://doi.org/10.1109/CVPR.2018.00068.

[135] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, IEEE Transactions on Image Processing 27 (1) (2018) 206–219, https://doi.org/10.1109/TIP.2017.2760518.

[136] H. Talebi, P. Milanfar, NIMA: Neural image assessment, IEEE Transactions on Image Processing 27 (8) (2018) 3998–4011, https://doi.org/10.1109/TIP.2018.2831899.

[137] E. Prashnani, H. Cai, Y. Mostofi, P. Sen, PieAPP: Perceptual image-error assessment through pairwise preference, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[138] L. Qu, G.W. Meyer, Perceptually guided polygon reduction, IEEE Transactions on Visualization and Computer Graphics 14 (5) (2008) 1015–1029, https://doi.org/10.1109/TVCG.2008.51.

[139] N. Menzel, M. Guthe, Towards perceptual simplification of models with arbitrary materials, Computer Graphics Forum 29 (2010).

[140] R.L. de Queiroz, P.A. Chou, Motion-compensated compression of dynamic voxelized point clouds, IEEE Transactions on Image Processing 26 (8) (2017) 3886–3895, https://doi.org/10.1109/TIP.2017.2707807.

[141] Z. Wang, E. Simoncelli, A. Bovik, Multiscale structural similarity for image quality assessment, in: The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003, vol. 2, 2003, pp. 1398–1402, https://doi.org/10.1109/ACSSC.2003.1292216.

[142] Z. He, G. Jiang, Z. Jiang, M. Yu, Towards a colored point cloud quality assessment method using colored texture and curvature projection, in: 2021 IEEE International Conference on Image Processing (ICIP), 2021, pp. 1444–1448, https://doi.org/10.1109/ICIP42928.2021.9506762.

[143] T. Chen, C. Long, H. Su, L. Chen, J. Chi, Z. Pan, H. Yang, Y. Liu, Layered projection-based quality assessment of 3D point clouds, IEEE Access 9 (2021) 88108–88120, https://doi.org/10.1109/ACCESS.2021.3087183.

[144] Z. Wang, Q. Li, Information content weighting for perceptual image quality assessment, IEEE Transactions on Image Processing 20 (5) (2011) 1185–1198, https://doi.org/10.1109/TIP.2010.2092435.
[145] Qi Liu, Hui Yuan, Honglei Su, Hao Liu, Yu Wang, Huan Yang, Junhui Hou, PQA-Net: Deep no reference point cloud quality assessment via multi-view projection, IEEE Transactions on Circuits and Systems for Video Technology 31 () (2021) 4645–4660, https://doi.org/10.1109/TCSVT.2021.3100282.
[146] W.-x. Tao, G.-y. Jiang, Z.-d. Jiang, M. Yu, Point Cloud Projection and Multi-Scale Feature Fusion Network Based Blind Quality Assessment for Colored Point Clouds, Association for Computing Machinery, New York, NY, USA, 2021, pp. 5266–5272.
[147] S.J. Daly, Visible differences predictor: an algorithm for the assessment of image fidelity, in: Human Vision, Visual Processing, and Digital Display III, vol. 1666, 1992, pp. 2–15.
[148] Q. Zhu, J. Zhao, Z. Du, Y. Zhang, Quantitative analysis of discrete 3D geometrical detail levels based on perceptual metric, Computers & Graphics 34 (1) (2010) 55–65, https://doi.org/10.1016/j.cag.2009.10.004.
[149] F. Caillaud, V. Vidal, F. Dupont, G. Lavoué, Progressive compression of arbitrary textured meshes, Computer Graphics Forum 35 (7) (2016) 475–484.
[150] P. Lindstrom, G. Turk, Image-driven simplification, ACM Transactions on Graphics 19 (3) (2000) 204–241, https://doi.org/10.1145/353981.353995.
[151] Ilyass Abouelaziz, Aladine Chetouani, Mohammed El Hassouni, Hocine Cherifi, A blind mesh visual quality assessment method based on convolutional neural network, in: Proc. IS&T Int'l. Symp. on Electronic Imaging: 3D Image Processing, Measurement (3DIPM), and Applications, 2018, pp. 423-1–423-5, https://doi.org/10.2352/ISSN.2470-1173.2018.18.3DIPM-423.
[152] I. Abouelaziz, A. Chetouani, M. El Hassouni, L.J. Latecki, H. Cherifi, No-reference mesh visual quality assessment via ensemble of convolutional neural networks and compact multi-linear pooling, Pattern Recognition 100 (2020) 107174.
[153] D. Plemenos, M. Benayada, Intelligent display in scene modelling. New techniques to automatically compute good views, in: Proceedings of the International Conference GraphiCon'96, 1996.
[154] X. Bonaventura, M. Feixas, M. Sbert, L. Chuang, C. Wallraven, A survey of viewpoint selection methods for polygonal models, Entropy 20 (5) (2018), https://doi.org/10.3390/e20050370, https://www.mdpi.com/1099-4300/20/5/370.
[155] K. Seshadrinathan, R. Soundararajan, A.C. Bovik, L.K. Cormack, A subjective study to evaluate video quality assessment algorithms, in: Human Vision and Electronic Imaging XV, vol. 7527, International Society for Optics and Photonics, 2010, p. 75270H.
[156] A. Ak, E. Zerman, S. Ling, P.L. Callet, A. Smolic, The effect of temporal sub-sampling on the accuracy of volumetric video quality assessment, in: 2021 Picture Coding Symposium (PCS), 2021, pp. 1–5, https://doi.org/10.1109/PCS50896.2021.9477449.