Towards Subjective Quality Assessment of Point Cloud Imaging in Augmented Reality

(Invited Paper)

Evangelos Alexiou, Evgeniy Upenik and Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015, Lausanne, Switzerland
Emails: firstname.lastname@epfl.ch

Abstract—**Recently, there has been an increased interest in capture, processing and rendering of visual content in form of point clouds. Among other challenges, subjective and objective quality assessments of point clouds are still open problems. Most proposed subjective quality evaluation methodologies are variants or extensions of counter parts from conventional approaches such as those proposed in various ITU-R and ITU-T recommendations. A key issue with point cloud content is that of rendering and display devices which are thoroughly different from those in other modalities in addition to novel applications which depart from traditional display devices. In this paper, we propose a radically different approach to point cloud subjective quality assessment for point cloud by making use of augmented reality head mounted displays. Beside description of the approach, we show examples of implementation of the proposed methodology and draw conclusions regarding its advantages and drawbacks. Finally, the proposed approach is used in assessing the performance of widely used objective metrics to compute quality of point cloud contents when they undergo various types of distortions such as corruption by noise, simplification and compression.**

I. Introduction

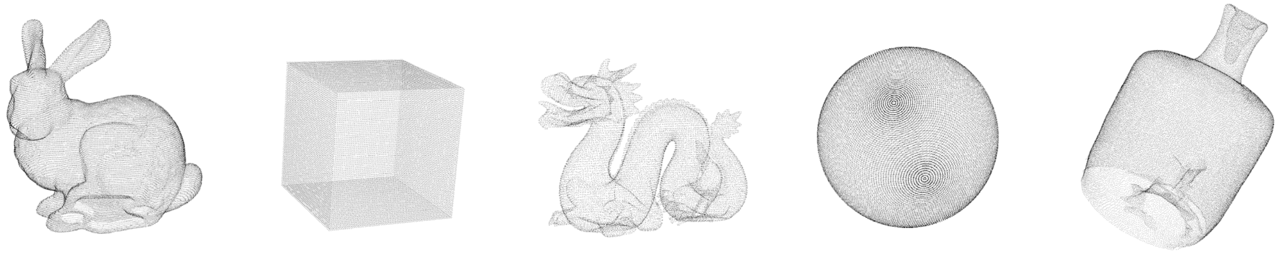
The current trend of adopting 3D technologies in imaging suggests that in the near future there will be a very substantial increase of new applications in virtual, augmented and mixed reality, digital elevation models, architecture, medical imaging and 3D printing, among others. A common and practical way for storage and rendering of 3D models in such applications is by using point clouds. It is also the default format used by acquisition devices that capture the depth of a scene (i.e., 3D scanners and depth sensors). A point cloud could be interpreted as a collection of three-dimensional points in space representing the external surface of an object. Each sample is defined by its position, which is obtained by the measured or reconstructed X, Y, and Z coordinates. Additional features can be associated with the coordinate data as well in order to provide further information, such as the point's color, normal or curvature.

Point cloud representation allows users to visualize image or video contents mimicking the perception of real-world scenes; in other terms, it is a viable solution to perceive 3D digital objects in a more immersive way. This feature can be exploited in related applications, such as in augmented reality. Augmented reality is a technology that extends the physical environment by introducing digital components into a person's senses. According to [1], three key requirements should be fulfilled: (a) combination of real-world and virtual assets, (b) interactivity in real-time, and (c) real and virtual imagery registered in 3D. Major improvements have been established to experience current implementations, after more than five decades of research and development in tracking algorithms, display and input devices, interaction techniques and usability. Nowadays, augmented reality is widespread in a number of applications in the areas of advertising, entertainment, education, medicine and mobile. Commercial devices, such as Microsoft Hololens, ODG Smart Glasses and Bridge Occipital, are just a few examples indicating the recent advances. However, this technology hasn't yet reached its full potential. Considering that research activities are currently ongoing and hardware specifications are continuously improving, in the upcoming years augmented reality will be part of our daily life, enabling emergence of exciting and new experiences.

Independently of the representation adopted in every type of application, the visual quality and the user experience are extremely important factors. The quality of a content is typically evaluated through objective or subjective assessments methodologies. In the first case computer algorithms designed to estimate the signal degradations are used. In the second case human subjects participate in experiments and rate the test contents. Subjective quality assessment is typically conducted based on ITU-R Recommendation BT.500-13 [2], where comparison methods, experimental designs, test methods, and evaluation procedures are explicitly defined. Subjective tests are expensive in terms of cost and time and, thus, objective quality assessment metrics are commonly used instead. However, objective metrics have to be properly calibrated to provide meaningful predictions of the subjective scores, which are considered as ground truth.

Until now, conventional approaches were followed in subjective quality assessment of point clouds. Traditional display devices were used and the subjective evaluation protocols did not exploit the full potentials of a richer representation. Using for instance head mounted displays, the user may interact with the content through a 6 degrees-of-freedom (6DOF)system in a more natural way. The real-time combination of real and virtual objects as well as the increasing level of immersiveness

978-1-5090-3649-3/17/$31.00 ©2017 IEEE



(a) Bunny (b) Cube (c) Dragon (d) Sphere (e) Vase

Fig. 1: Selected contents

may affect the perceptual quality and subjective scores may differ.

In this paper we propose the use of head mounted displays in augmented reality scenarios for subjective quality assessment of point cloud geometry. Furthermore, the subjective scores are correlated with state-of-the-art objective metrics. During the experiments, subjects visualize, interact and rate the level of impairment of point cloud geometry, after introducing typical degradations such as noise and compression-like distortions. Our results show that current objective metrics perform quite well in presence of noise, but fail to accurately predict the perceptual quality for every type of content in presence of compression-like artifacts.

II. Related Work

A limited amount of work on subjective assessment of point cloud data has been reported in the literature. In [3], a 3D tele-immersive system was proposed and users represented by avatars (i.e., 3D point clouds captured by multiple Microsoft Kinect sensors) were able to interact in a virtual room resulting in a complex scenario. In [4], the subjective assessment of point clouds was performed for different resolutions and values of geometric and color noise. The uniform noise that is considered, though, does not correspond to a realistic model of noise for point clouds, neither for geometric nor for color degradations. Furthermore, in both cases, the point-to-plane metric is not taken under consideration and the correlation between objective and subjective metrics is not reported. In [5], quality evaluation of point cloud denoising algorithms is proposed. In this study, impulse noise was initially introduced to simulate outlier errors. After applying the radius outlier removal algorithm, Gaussian noise was added to the processed models to mimic sensor imprecisions. Two denoising algorithms were then applied to the degraded content and subjectively assessed in a passive way. To visualize the content, the Poisson surface reconstruction method was used and the resulted 3D object was captured from different viewpoints with a specific pattern to form a video. The results were correlated with several state-of-the-art objective metrics. However, as it is clearly stated, the scope of this paper was to assess denoising algorithms for point clouds rather than quality of the content.

In [6], an interactive subjective evaluation of point cloud geometry is proposed. The participants were asked to assess the level of impairment, while they were able to visualize and interact with both the original and the processed contents simultaneously. However, a typical flat screen was used and the interaction between the user and the content was not natural.

III. Subjective Experiments

This section reports how the subjective quality evaluation experiments were designed. Specifically, the creation of the contents, the adopted distortions, the equipment and the testing environment are described.

A. *Selection of contents*

In this experiment, subjective quality assessment of point cloud geometry of five contents is performed. In order to assess only the geometrical distortions, no color attribute was assigned to the points. For this reason simple objects were selected, since it would be difficult for complex scenes to be distinguishable in absence of color. Furthermore, to normalize the impact of distortions (specially for noise), they were scaled to be fitted in a bounding box of size 1. Different acquisition techniques were considered in order to increase the diversity of the structure of the point cloud contents. In particular, *bunny* and *dragon* were selected from the Stanford 3D Scanning Repository[1] to represent widely used contents with reduced noise after post-processing. *Cube* and *sphere* are artificially generated and represent synthetic contents with perfect geometry. Finally, *vase* is a model captured by Intel RealSense R200 and constitutes a representative point cloud that can be acquired from low-cost consumer market device. In Figure 1 the selected contents are displayed, while in Table I their corresponding number of points is provided.

TABLE I: Number of points per content

Contents:	Bunny	Cube	Dragon	Sphere	Vase
Points:	35947	30246	22998	30135	36022

It should be mentioned that in order to avoid performance issues related to the equipment used for the experiments, a sparse version of the *dragon* was used (i.e., namely, *dragon_vrip_res3*), and the initially captured *vase* was downsampled. For the former case, a minimal distance between two points was set to a specific value. Using CloudCompare[2],

[1] http://graphics.stanford.edu/data/3Dscanrep/

[2] http://www.cloudcompare.org

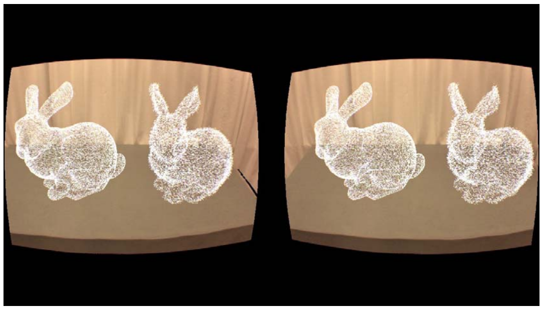


Fig. 2: Rendering application screen shot showing the reference on the left, and a stimulus impaired with Gaussian noise of $\sigma = 0.008$ on the right.

a downsampled version of the original cloud was generated by ensuring that no point in the output cloud is closer to another point than the specified value. Furthermore, no displacements in the original coordinates of the points were introduced, maintaining the default irregular structure of this content.

B. Degradation Types

In this study, two different types of geometrical degradations are assessed. In the first case noise is introduced, while in the second case the processed content consists of a sparser and more regular version of the original point cloud, which is obtained after applying an octree structure with appropriate resolution values. Thus, in the latter case, the processed content is subject to points displacements and points removals. The values of parameters to define both types of distortions were selected to represent subjective quality covering a wide range from lowest to highest levels.

1) Noise: Gaussian noise models position errors due to imperfections in acquisition from depth sensors. This model is widely used in the literature. In our case, the noise affects the position of all points of the point cloud and its level is determined by a target standard deviation (i.e., $\sigma = \{0.0005, 0.002, 0.008, 0.016\}$).

2) Octree-pruning: Octree structure is extensively adopted in point cloud compression algorithms as it enables organized representation of points, which is further exploited to reduce the size of data needed for content reconstruction. This regular representation, though, leads to visible artifacts in the form of structured distortions. Octree-pruning can be obtained by setting a desirable octree resolution; this defines the size of leaf nodes. The content is included in a bounding box and, in each level, each cube is sub-divided into 8 smaller and equally sized cells. A point can be appended only in leaf nodes and all points within a leaf node are collectively represented by the center of that node leading to both points removals and points displacements, limited by the diagonal of the leaf node divided by two. This way by increasing the octree resolution, the number of remaining points decreases. Octree-pruning could be interpreted as an instance of progressive decoding procedure, as it creates similar distortions. In our case, octree resolution values are selected for each content in order to obtain target percentages (p), with respect to the original number of points, with an acceptable deviation of $\pm 2\%$ (i.e., $p = \{30\%, 50\%, 70\%, 90\%\}$).

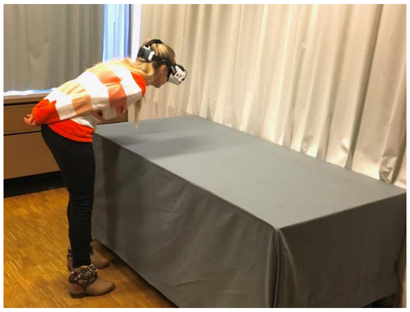


Fig. 3: Participant observing the stimuli in augmented reality using head-mounted display.

C. Environment and Equipment

The experiments were conducted in a test laboratory which fulfills the recommendations for subjective evaluation of visual data issued by ITU [2]. A test table covered by a medium gray tissue was installed in the room. The subjects were observing the stimuli in augmented reality environment provided by a hardware-software system developed by the authors. The testbed is based on Occipital Bridge AR headset[3]. iPhone 6S was used as a screen providing the resolution of 326 pixels per inch. Occipital Bridge software development kit libraries allow rendering of a real world scene captured by the phone's camera with an attached wide angle lens of 120 degree field of view. The point cloud objects are rendered on top of the scene by means of SceneKit library. Each point is represented by an atomic triangle of a size significantly smaller than the object dimensions. Thus, these triangles are perceived as points by the viewer. Each point is of white color with saturated luminosity. The brightness values of the points and the test table surface were measured on the phone's screen with a luminosity sensor[4] providing the values 595.28 and 38.91 nits, respectively. The assessed objects were placed in fixed locations on the test table (Fig. 2) in augmented reality. The subjects were initially asked to stand in front of the test table at the distance of 1 meter and they were free to change their position after the beginning of each evaluation session.

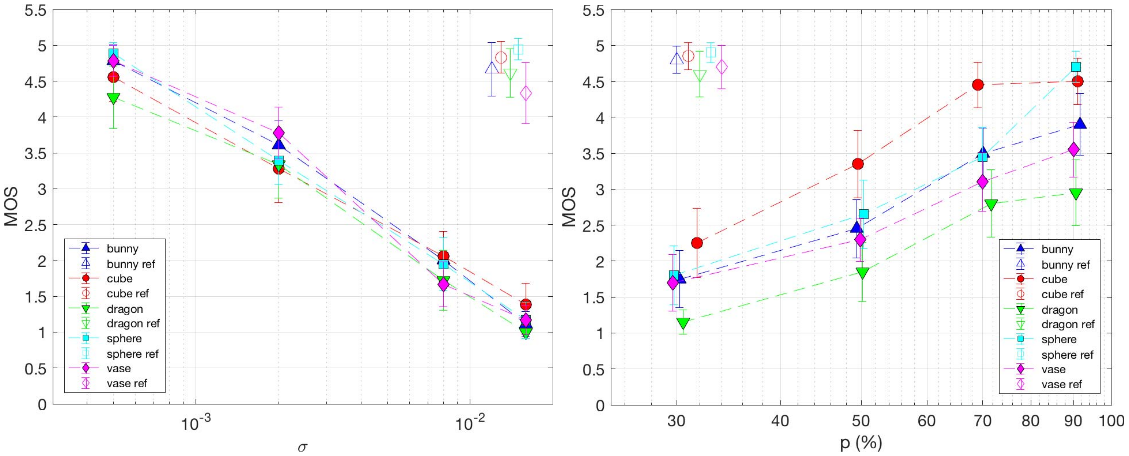
D. Subjective Evaluation Methodology

The double-stimulus impairment (DSIS) with 5-scale rating was selected for its high accuracy and reliability in constructing a scale of perceptual references. Essentially, as this is the first attempt of interactive assessment of point clouds, the exact impact of each available evaluation method is not known and, thus, a basic and widely adopted approach was decided to be used. The original and the processed stimuli were displayed simultaneously, resulting in a side-by-side visualization.

In order to reduce contextual effects, the position of the reference while remaining fixed across each session for every subject, changed randomly for different subjects. Furthermore,

[3] https://bridge.occipital.com/

[4] X-Rite i1 Display Pro - http://www.xrite.com/



(a) MOS vs standard deviation of Gaussian noise

(b) MOS vs percentage of points used in octree structure

Fig. 4: Subjective scores for each type of degradation

the order of the observed pairs was randomized per session and subject, ensuring that the same content had not been shown consecutively. The subjects were free to interact with the content by walking around the scene (e.g. coming closer, changing the point of view, etc.) during the evaluation procedure (Fig. 3). There was no time limitation. After inspecting each pair, the subjects were listening to the rating scale before providing their scores orally.

Since two different types of degradations were assessed, the evaluation procedure was split in two different sessions. Each session was launched after a training phase, where the subjects were informed about the general characteristics of the type of noise they would assess and got familiarized with the interaction part. They were specifically instructed to rate the level of impairment in terms of how annoying the degraded stimuli is for them with respect to the reference.

For each session, 5 contents and 4 degradation values were used along with a hidden reference resulting in 25 stimuli per session. A total of 24 naive subjects (14 males and 10 females) participated in the experiments; 18 of them were involved in both sessions while 6 participated in just one, leading to 21 scores per stimulus. The age was ranging from 25 to 32 with an average of 27.66 and a median of 28 years of age.

IV. Statistical Analysis

In this section an overview of the state-of-the-art objective metrics is provided. A description of the statistics obtained from the subjective scores and the methodology to compute the performance indexes of the objective metrics are also reported.

A. *Subjective Quality Assessment Methodology*

For each session, outlier detection and removal is performed on the results based on Recommendation ITU-T P.1401 [7]. Considering the first session, three outliers were detected, whereas one outlier was found in the second. Thus, 18 out of the 21 scores and 20 out of the 21 scores per stimulus were used for the first (i.e., Gaussian noise) and second (i.e., Octree-pruning) types of distortions, respectively. After applying outlier detection, the mean opinion score (MOS) was computed for each degraded content, along with the corresponding 95% confidence intervals assuming a Student's t-distribution.

B. *Objective Quality Metrics*

In objective evaluation of point clouds, similarity is the key factor to assess the geometry of a processed 3D content. It works on the basic principle of getting the quantitative distance between the processed and the original content. The state-of-the-art objective metrics for geometric distortions can be classified as point-to-point (p2point), point-to-plane (p2plane) or point-to-mesh (p2mesh) [8]–[10]. p2mesh distances heavily depend on the surface reconstruction technique that is being used for the mesh and, hence, they can be considered as suboptimal. The p2point error is calculated by connecting each point of the point cloud under evaluation to the closest point of the reference. The p2plane error measures the projected error along the normal of the closest point of the reference point cloud [10]. Geometric errors between point clouds can be estimated either using the root mean square (RMS) difference or the Hausdorff[5] distance for both p2point and p2plane cases. Commonly, the symmetric distance is used; that is, obtained by setting both the original and the processed content as reference and estimate both errors. Then, the maximum value is considered [9]. However, such absolute values of error fail to assess the difference between differently scaled contents. For this purpose, Peak-to-Signal Noise Ratio (PSNR) ratio is proposed. In the literature, it is defined as the ratio of the squared maximum distance of the nearest neighbours, or the squared distance of the diagonal of the bounding box divided by the squared error value (i.e., squared RMS or squared Hausdorff). In this study, the first definition of PSNR is adopted. Finally, all the possible combinations of the distances and metrics are considered, leading to a total of 8 different objective metrics.

[5] The Hausdorff distance is defined as the greatest of all the distances from a point in one set to the closest point in the other set.

TABLE II: Performance indexes for the different metrics

Metric	Gaussian noise				Octree-pruning			
	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
$\text{p2point}_{\text{RMS}}$	0.9890	0.9383	0.2085	0.15	0.5124	0.4286	0.8750	0.65
$\text{p2plane}_{\text{RMS}}$	0.9888	0.9353	0.2099	0.10	0.4854	0.4887	0.8908	0.55
$\text{p2point}_{\text{Hausdorff}}$	0.9904	0.9293	0.1943	0.10	0.5451	0.5297	0.8542	0.55
$\text{p2plane}_{\text{Hausdorff}}$	0.9896	0.9398	0.2023	0.10	0.5306	0.4406	0.8636	0.55
PSNR - $\text{p2point}_{\text{RMS}}$	0.9871	0.9526	0.2255	0.25	0.5632	0.5263	0.8420	0.55
PSNR - $\text{p2plane}_{\text{RMS}}$	0.9905	0.9526	0.1941	0.25	0.5651	0.5338	0.8406	0.55
PSNR - $\text{p2point}_{\text{Hausdorff}}$	0.9911	0.9503	0.1880	0.20	0.6340	0.6586	0.7880	0.45
PSNR - $\text{p2plane}_{\text{Hausdorff}}$	0.9901	0.9526	0.1978	0.30	0.5808	0.5549	0.8294	0.55

C. Performance Indexes

Subjective MOS are used as the ground truth in order to benchmark the objective metrics. The result of execution of a particular objective metric is a Point cloud Quality Rating (PQR). A predicted MOS, denoted as MOS_P, is estimated by applying a fitting function to each [PQR, MOS] pair, with respect to the degraded stimuli. According to Recommendation ITU-T P.1401 [7], the following properties of the PQR are considered: linearity, monotonicity, accuracy and consistency, by computing the Pearson linear correlation coefficient (PCC), the Spearman rank order correlation coefficient (SROCC), the root-mean-square error (RMSE) and the outlier ratio (OR) between MOS and MOS_P, respectively. Linear, logistic and cubic fittings were tested and it was found that the latter provides the best results. Thus, the cubic fitting is adopted to demonstrate our results.

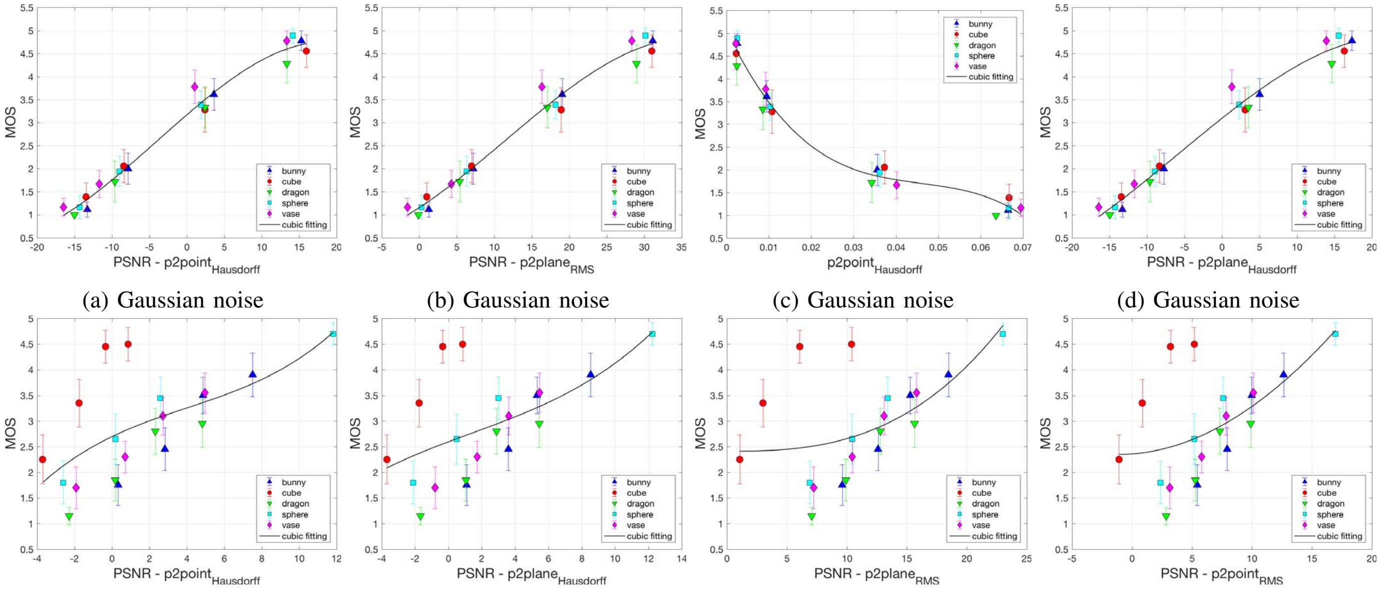
V. Results and Discussion

In Figure 4 the MOS along with the confidence intervals against the degradations values are depicted. The markers with faces indicate the scores for the distorted versions of the original point clouds, while the markers without faces (i.e., at the top-right of Figure 4a and top-left of Figure 4b) correspond to the scores of the hidden references. It can be observed that as the standard deviation of the Gaussian noise is increasing, the MOS is decreasing following a logarithmic trend. The subjects seem to be able to recognize easily the amount of noise introduced, independently of the displayed content. The DSIS methodology adopted, also, assists to obtain such results, since the subjects are always aware of the reference content and their ratings are based on relative geometrical differences.

Conversely, when the contents are subject to compression-like distortions, the underlying surface and shape of the content seem to play a significant role. For instance, *cube* is rated remarkably higher than any other content, except for *sphere* for $p = 90\%$. Apparently, the more complex is the model, in terms of curvature, the lower the MOS are. For example *dragon*, which is the most complex object, is notably under-rated. Any removal of points for this particular object has higher impact, and even for $p = 90\%$ the MOS is much lower than the MOS of the hidden reference. Another reason for *dragon*'s scores is its geometry. As mentioned in Section III-A, the contents were scaled in a range between $[0, 1]$. The shape of the *dragon* and, specifically, the ratio between height and length is such that it does not fill the bounding box entirely, so the content looks remarkably smaller than the others. Since the subjects mostly kept a fixed distance during the evaluation procedure, perceiving one object as smaller than the others may have affected its rating. *Vase*'s irregular structure is transformed to a regular representation after octree-pruning. As subjects tend to rate based on relative differences, the MOS of the *vase* is systematically lower then the MOS of *bunny* and *sphere*, which are more regular contents. Finally, as *sphere* is artificially generated, the density of points in poles is much higher. For $p = 90\%$, no remarkable impairments occur in the remaining surface and, thus, it is rated similarly to the hidden reference.

In Table II, the performance indexes for the current state-of-the-art objective metrics are presented. Furthermore, in Figure 5 the scatter plots of subjective scores against the 4 most efficient objective metrics are displayed for noise and octree-pruning. As it can be observed, our results show strong correlation between objective metrics and subjective scores in the presence of Gaussian noise. Considering that the objective metrics are full-referenced and all of them are based on Euclidean distances of neighbouring points between the original and the processed contents, by increasing the standard deviation of noise, the obtained results worsen. On the other hand, subjects were able to visualize both reference and the degraded point clouds side-by-side and, thus, they could relatively easily identify the level of discrepancies. This explains why the results show such a strong correlation.

On the contrary, the correlation between subjective and objective scores in the presence of compression-like distortions is poor. Despite the fact that the level of perceptual impairment is reflected in the objective scores for contents with curved surfaces, this is not the case for objects with planar surfaces. As the number of points decrease, less details and more rough representations of curved edges are observed, which lead subjects to rate the processed content with lower scores. Thus, the objective results, which are based on distances of the closest points between the reference and the distorted contents, are coarsely aligned to the subjective scores. However, this is not the case for point clouds that consist of planar surfaces. Furthermore, the octree structure, by default, arranges the points of the processed object in a structured and equally spaced way. Thus, the structured loss is not perceived as truly annoying, since it does not affect the underlying structure of the object in the case of the *cube*. Additionally, subjects tend to



(a) Gaussian noise (b) Gaussian noise (c) Gaussian noise (d) Gaussian noise

(e) Octree-pruning (f) Octree-pruning (g) Octree-pruning (h) Octree-pruning

Fig. 5: Subjective vs objective results

rate higher when the structure of the point cloud is regular, and in the case of the *cube* the regularity is maintained. It should also be mentioned that by removing the scores for *cube*, the correlation of the subjective scores and objective metrics is significantly improved, with p2point metric using Hausdorff distance achieving the best performance (i.e., PCC $= 0.9359$).

Finally, based on observations extracted during the experimental procedure, there were very few cases of subjects that were feeling confident with the interactivity part, as most of them preferred to stay static. The level of interaction and the viewing position are important factors, and in order to compensate their impact on the MOS and confidence intervals, we would suggest to use more than 15 subjects as proposed in the case of quality assessment for conventional content. Furthermore, subjects tend to rate objects based on the number of points and, in general, they prefer regular representations. For instance, in the case of *vase* for $p = 90\%$, a few subjects asked why there is no option to rate the processed content higher than the reference. Thus, it would be interesting to perform subjective tests using Absolute Category Rating or Pair Comparison to get further insights.

VI. Conclusions

In this paper we propose the use of augmented reality in order to subjectively evaluate the quality of point cloud geometry. In addition, state-of-the-art objective metrics were correlated with the subjective scores. The statistical analysis shows that the current metrics perform well when Gaussian noise is introduced. However, in the presence of compression-like artifacts the performance is lesser for every type of content, leading to a conclusion that the performance is content dependent. Our results show that there is a need for better objective metrics that can more accurately predict all practical types of distortions for a wide variety of contents.

Acknowledgment

This paper reports a research performed in the framework of a collaborative project Random Access Video Encoding (RAVE) with Teleport SA, funded by The Swiss Commission for Technology and Innovation (CTI) (grant 18875.1PFES-ES) and also in the framework of ImmersiaTV under the European Unions Horizon 2020 research and innovation programme (grant agreement no. 688619) funded by Swiss State Secretariat for Education, Research and Innovation SERI.

References

[1] R. T. Azuma, “A Survey of Augmented Reality,” *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, Aug. 1997.

[2] ITU-R BT.500-13, “Methodology for the subjective assessment of the quality of television pictures,” International Telecommunication Union, January 2012.

[3] R. Mekuria, K. Blom, and P. Cesar, “Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828–842, Apr. 2017.

[4] J. Zhang, W. Huang, X. Zhu, and J. N. Hwang, “A subjective quality evaluation for 3d point cloud models,” in *International Conference on Audio, Language and Image Processing*, 2014, pp. 827–831.

[5] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, “Subjective and objective quality evaluation of 3d point cloud denoising algorithms,” ISO/IEC JTC m75024, Sydney, Australia, March 2017.

[6] E. Alexiou and T. Ebrahimi, “On subjective and objective quality evaluation of point cloud geometry,” in *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*, 2017.

[7] ITU-T P.1401, “Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models,” International Telecommunication Union, July 2012.

[8] P. Cignoni, C. Rocchini, and R. Scopigno, “Metro: Measuring error on simplified surfaces,” *Computer Graphics Forum*, vol. 17, no. 2, pp. 167–174, June 1998.

[9] R. Mekuria, Z. Li, C. Tulvan, and P. Chou, “Evaluation criteria for point cloud compression,” ISO/IEC MPEG n16332, Geneva, Switzerland, February 2016.

[10] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, “Evaluation metrics for point cloud compression,” ISO/IEC JTC m74008, Geneva, Switzerland, January 2017.