Subjective and Objective Quality Assessment for Volumetric Video Compression

Emin Zerman, Pan Gao, Cagri Ozcinar, Aljosa Smolic
V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

Abstract

Volumetric video is becoming easier to capture and display with the recent technical developments in the acquisition, and display technologies. Using point clouds is a popular way to represent volumetric video for augmented or virtual reality applications. This representation, however, requires a large number of points to achieve a high quality of experience and needs compression before storage and transmission. In this paper, we study the subjective and objective quality assessment results for volumetric video compression, using a state-of-the-art compression algorithm: MPEG Point Cloud Compression Test Model Category 2 (TMC2). We conduct subjective experiments to find the perceptual impacts on compressed volumetric video with different quantization parameters and point counts. Additionally, we find the relationship between the state-of-the-art objective quality metrics and the acquired subjective quality assessment results. To the best of our knowledge, this study is the first to consider TMC2 compression for volumetric video represented as coloured point clouds and study its effects on the perceived quality. The results show that the effect of input point counts for TMC2 compression is not meaningful, and some geometry distortion metrics disagree with the perceived quality. The developed database is publicly available to promote the study of volumetric video compression.

Introduction

Capturing and displaying volumetric video are becoming easier with the developing technology for augmented reality (AR), mixed reality (MR), and virtual reality (VR) applications. Recent technological developments made the acquisition of volumetric video feasible [1–3]. The acquired volumetric video can easily be used within various AR and VR applications using the head-mounted displays (HMD). Thus, within both industry and the scientific community, the interest to the effective acquisition and display of volumetric video is increasing [3–5].

A popular way of volumetric video representation is using point clouds (PC). As PCs are getting used more commonly [4,6], the interest and need for creating a content delivery chain for PCs increase. In this context, both compression and quality assessment are crucial aspects to improve and spread the use of PCs and volumetric videos. Although there are both subjective [6–14] and objective [15–17] studies on the quality assessment of PCs, the quality assessment of coloured PC compression for acquired volumetric video is still very rare. In fact, to the best of our knowledge, this study is the first to consider TMC2 compression for volumetric video represented as coloured PCs and study its effects on the perceived quality.

In this work, we subjectively and objectively evaluate the quality perception for the volumetric video compression scenario using a state-of-the-art PC compression method: MPEG Point Cloud Compression Test Model Category 2 [18] (denoted as TMC2). The contributions of this paper are threefold: *(i)* The effects of the TMC2 compression method on the perceived quality were analysed for the first time through subjective experiments, *(ii)* the relationship between the human observers' perception and the prediction performance of the state-of-the-art objective quality metrics was studied, and *(iii)* an initial subjective quality database was created to stimulate further research in this subject.

Detailed information on the related work, the volumetric video quality database created, subjective experiments conducted, objective quality estimation, and results is presented in the following sections.

Related Work

A 3D model can be represented using PCs without the need of connectivity information and a texture atlas. Instead, the colour information can be provided with each point. However, for proper visualisation, the number of points should be high, and this increases the size of PCs. Compression is necessary to store and transmit the PCs. Three compression approaches are mainly considered for point cloud compression (PCC): octree-based compression [4, 19], graph-based compression [10], and video-based compression [18].

A few point-based objective measures were proposed to evaluate the geometry distortion in PCC: point-to-point [15], point-to-plane [16], and plane-to-plane [17]. Although these metrics are the state-of-the-art and used in many recent studies [6, 9–13] and in MPEG standardisation activities [4,18], the performance of these metrics was found to be not sufficient to predict the visual quality for different types of contents [12, 13].

To evaluate these objective quality metrics and to understand the effect of visualisation, many subjective quality assessment studies were conducted. Most of these studies concentrate on computer-generated PCs without colour information [6, 8, 9, 12–14], and all of these studies used Gaussian noise or octree pruning as distortions. Some of these studies compared subjective test methodologies [9] and different visualisation strategies [14] which did not yield statistically significant difference. In another study, it is found that visualising PCs as raw PCs and reconstructed surfaces yield statistically significant difference [12].

Coloured PCs were considered only in a few studies. Zhang et al. [7] considered only Gaussian noise either in the location or colour of the points. Mekuria et al. [4] conducted a subjective evaluation for the validation of the proposed codec. Javaheri et al. [10] considered an octree-based and a graph-based compression method and compared geometry-only point-based objective

Figure 1. *Sample rendered images for PCs at different quality levels: uncompressed reference (leftmost) and compressed PCs where i=1 (centre-left), i=2 (centre), i=3 (centre-right), and i=4 (rightmost).*

Figure 2. *Sample 2D rendered versions of the two selected volumetric videos; (left) 'Matis' and (right) 'Rafa'.*

quality metrics. Torlig et al. [11] also considers an octree-based method and evaluates the quality of compressed PCs both subjectively and objectively. It is found that projection-based objective quality metrics (i.e. traditional image quality metrics) are better correlated to human viewers' scores compared to the point-based objective PC quality metrics. Although these studies focus on octree-based or graph-based compression methods, to the best of our knowledge, video-based compression methods have not been studied yet.

In this work, we evaluate the video-based TMC2 PCC method both subjectively and objectively.

Volumetric Video Quality Database

To create a volumetric video quality database, we use two volumetric videos generated by the V-SENSE research group using an affordable set-up with 12 synchronised RGB cameras [3]: Two different people playing with a football as our uncompressed source videos (named 'Matis' and 'Rafa'). Sample 2D rendered versions of these two selected models are presented in Fig. 2. These PCs were selected to have 4 different sampling frequencies (i.e. point counts): $p \in \{62K, 127K, 250K, 495K\}$ to understand the effect of input point counts on the compression, the subjective quality perception, and objective quality estimation.

The PCs were compressed using the state-of-the-art MPEG Point Cloud Compression Test Model Category 2 (TMC2) method [18]. This compression method is based on the H.265/HEVC standard [20]. The PC location and colour information are mapped to image pixels and encoded. To see the effect of compression parameters, 4 different quantization parameters (QP) were selected. TMC2 has two different quantization parameters: geometry QP (gQP) and texture QP (tQP). During initial trials, the effect of geometry QP was found to be harsher compared to that of texture. In order to keep the reduction in the perceived quality balanced, a ratio between gQP and tQP was found and was kept the same for these 4 quality levels. The following gQP and tQP values were selected in a pilot study done: $(gQP, tQP) \in \{(17,20),(30,35),(37,43),(41,48)\}$. The PCs were then encoded with these gQP_i and tQP_i values for each quality level $i \in [1,4]$, using TMC2. In total, 32 different processed video sequences (PVS) were generated for 2 contents $\times$ 4 point counts $\times$ 4 QPs. Together with the subjective data collected, this volumetric video quality database is made publicly available[1].

Visualisation for subjective experiment

For visualisation of PCs, each point was replaced with small elliptical planes (i.e. splats) to remain true to the data representation of PCs. The size of these splats was arranged so that there would be as little holes as possible (ideally none – see the centre-right image in Fig. 3), without causing the model to look swollen. Sample 2D rendered images are presented in Fig. 3 which show the differences in visualisation with respect to splat size. For this purpose, the PCs were loaded to Unity (version 2017.4.1f1), and the scene was rendered and recorded using Unity Recorder tool. The volumetric videos were placed on the origin and rotated on their z-axis to avoid inter-subject variation due to interaction. The objects' heights were set to 0.5 units, and a camera with perspective projection and 60-degrees field of view was placed 1.143 units away. The videos were comprised of 198 frames, which took 6.6 seconds in 30 frames/second.

Subjective Quality Assessment

As 'quality' is directly related to human perception, subjective quality assessment is the best method to assess multimedia quality. Using single or double stimulus direct rating methods [21] is the most common way to conduct subjective tests. However, very small differences may not be captured by these tests. Pairwise comparisons (PWC) method is easier for subjects to decide; however, it is very hard to obtain accurate and meaningful results when the preference ratios approach 100% (or 0%) [22]. Hence, PWC experiments are not suitable for large quality differences.

In this study, we conduct two subjective experiments to analyse both large and small differences among our stimuli. To cap-

[1] `http://v-sense.scss.tcd.ie/research/fvv/quality-assessment-for-fvv-compression/`

Figure 3. *Four different sample rendered images showing the effect of point size on rendered images. Splat size increases from left to right. The holes are visible in the leftmost and centre-left images, and the rightmost image looks swollen, as the text on chest also changes shape. Thus, the splat size corresponding centre-right image was selected.*

ture the larger differences introduced by compression a rating experiment was conducted with a double stimulus impairment scale (DSIS) methodology [23]. To capture smaller differences which can be induced by input point count parameter, a second experiment was conducted with the pairwise comparisons (PWC) methodology [21]. All the parameters were kept the same such as experiment set-up, stimuli, and also the subjects.

The experiment set-up and all the experiment variables were adjusted according to the ITU Recommendations [21,23]. The experiments were conducted in a special dark room, and the distance between the display and the viewers was fixed (i.e. $3H \approx 50$ cm). In total, 19 volunteers (15 M and 4 F) participated in the experiments, with the mean age was 31.0 (std. 0.23). Two experiments were conducted in two different sessions, and the same subjects were asked to attend both experiments. The experiment interface was prepared on Matlab using Psychtoolbox[2] [24].

Except for the experiment set-up, some other parameters were common in these two experiments. A training session was held before the experiments to familiarise the subjects with the experiment, with another volumetric video which was not presented during the test. The presentation was done in a passive manner without user interaction. For this purpose, the videos were rendered in Unity, as described in ***'Volumetric Video Quality Database'*** section above. The videos (6.6 secs) were shown twice to the viewers to make sure that they were able to inspect the videos adequately. The subjects were able to skip the second loop if they desired to. After the presentation, the subjects were presented with the voting screen and asked to vote.

Double stimulus impairment scale (DSIS) experiment

In the DSIS [23] experiment, both the reference and distorted (i.e. compressed) volumetric videos were presented side-by-side fashion. In order to avoid any bias with the replacement of the reference video, the reference video was placed on the right for half of the subjects and on the left for the other half.

After training, the subjects were presented with the stimuli, and they were asked to vote using the mouse. Subjects were asked to 'rate the quality impairment of compressed video with respect to the reference'. A continuous scale was used for voting, where the users were able to vote between [0, 100] values (100 corresponding to *'Imperceptible'* level), and the impairment adjectives *'Imperceptible'*, *'Perceptible'*, *'Slightly annoying'*, *'Annoying'*, and *'Very annoying'* were printed under the scale to help

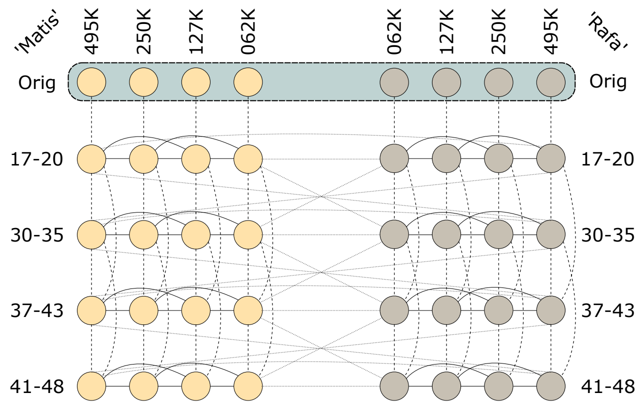


Figure 4. *The selected pairs for the pairwise comparisons experiment. The dashed lines indicate the comparisons between different quality levels, solid lines indicate comparisons among the videos with different source point counts, and dotted lines indicate cross content comparisons. For the psychometric scaling, all of the original videos were considered as one "original" or "reference" case where it is the ideal state of the video.*

the subjects. There was no time limit for voting, and subjects responded in 2.4 seconds on average for each stimulus. The DSIS experiments took 10 minutes on average, including the training.

Pair-wise comparisons (PWC) experiment

PWC [21] experiment was conducted to capture smaller differences and understand the effect of input point counts on the quality. Two different videos were shown to the subjects side-by-side fashion. Pairs to be compared were selected manually, instead of using a full-design PWC experiment. Cross-content pairs were found to increase accuracy and unify the scale of the scaling results [25]. Therefore, they were included to obtain more accurate psychometric scaling results and to be able to compare the scaling results of different contents. The selected pairs include *(i)* within-content within-point-count across-quality, *(ii)* within-content across-point-count within-quality, *(iii)* cross-content within-point count within-quality, *(iv)* cross-content within-point count across-quality. Additionally, the highest quality compressed videos (i.e. $QP = (17, 20)$) are also compared to their original uncompressed PCs. Fig. 4 shows all of the selected pairs.

Similar to the DSIS experiment, the subjects were presented with the stimuli after a brief training, and they were asked to vote using the keyboard. Subjects were asked to 'select the stimulus they preferred over the other', and they were able to use left or

[2]http://psychtoolbox.org/

right arrow of the keyboard for that purpose. There was no time limit for voting, and subjects responded in 0.6 seconds on average for each pair. The PWC experiments took 23 minutes on average, including the training.

Objective Quality Assessment

Objective quality evaluation for a media is essentially based on the data type used. For the objective evaluation of PCs, the only data we have is the point location. We may have additional colour information if the PC is coloured. Considering the availability of point location and colour information, either point-based metrics [15–17] can be used, or the PCs can be rendered as images and traditional image quality metrics can be used. The latter is also known as projection-based quality evaluation [11]. In this study, we only consider point-based metrics.

Three different point-based quality metrics were commonly used in the recent studies [6, 8–14]: point-to-point (po2point) [15], point-to-plane (po2plane) [16], and plane-to-plane (pl2plane) [17]. These metrics consider only the geometry errors, and they are computed using either finding *(i)* the root mean square (RMS) distance, *(ii)* mean square error (MSE) or *(iii)* Hausdorff distance. For colour errors, the difference in terms of MSE and PSNR is found between a point in the distorted PC and the nearest corresponding point in the reference cloud. The colour differences are computed for Y, U, and V channels. To understand the effect of compression on colour, we use the following weighted averaging method [26], as it is also done in a recent PC quality assessment work [11]:

$$PSNR_{YUV} = (6 \times PSNR_Y + PSNR_U + PSNR_V)/8 \qquad (1)$$

To compute the objective scores, we use version 0.09 of the software described in [27].

Experimental Results

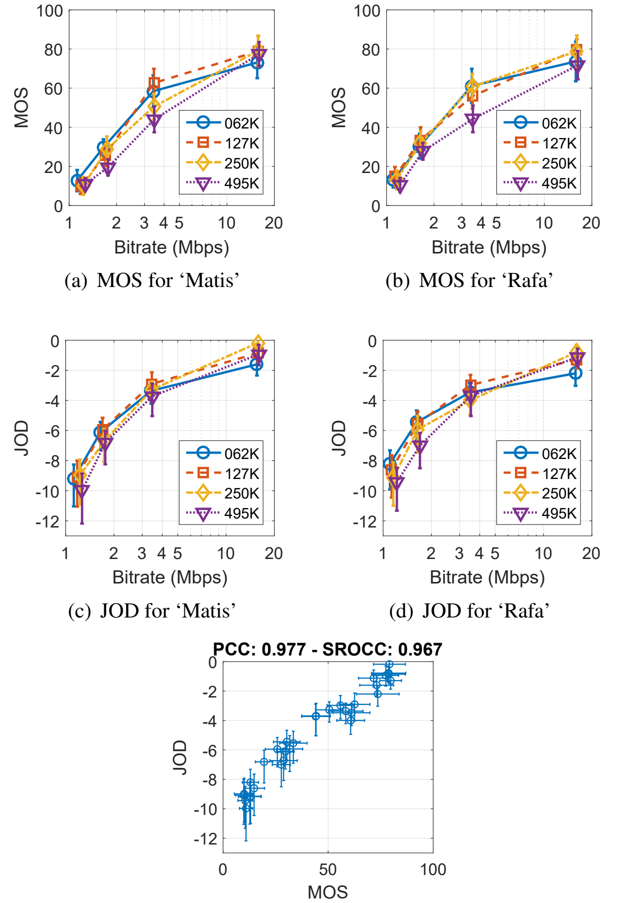
Data Analysis for Subjective Experiment Results

After their collection, the subjective quality scores should be checked for outliers to avoid noisy results. The analysis and outlier detection methods differ by the methodology used. Therefore, we used two different outlier detection methods for DSIS and PWC experiments, as explained below. Following the outlier detection and removal step, the subjective experiment results were analysed using a number of statistical analysis methods.

For DSIS experiment, the outlier detection step was done as it is recommended in ITU-R Recommendation BT.500 [23], and no outliers were found. Then, the results were analysed by finding the mean opinion scores (MOS) and confidence intervals (CI) as described in ITU-T Recommendation P.1401 [28].

For PWC experiment, the outlier removal method proposed in the work of Perez-Ortiz and Mantiuk [22] was used. Only one outlier was found, whose votes were removed. The results of the PWC experiment were then analysed using both a binomial test and using psychometric scaling results. For psychometric scaling, we used an open source scaling code[3] which assumes Thurstone Case V and estimates quality scores via maximum-likelihood estimation [22]. The resulting quality estimates are named as just objectionable differences (JOD), and their confidence intervals were found via bootstrapping.

[3] `http://github.com/mantiuk/pwcmp`



(a) MOS for 'Matis' (b) MOS for 'Rafa'

(c) JOD for 'Matis' (d) JOD for 'Rafa'

(e) MOS vs. JOD

Figure 5. *Plots for the subjective quality assessment results, for (a, b) MOS and (c, d) JOD. Please note that the x-axes of subplots a, b, c, and d are in logarithmic scale for a better visualisation. Additionally, a plot for MOS vs. JOD (e) was included for comparison.*

Evaluation of TMC2 compression

The evaluation of TMC2 compression and some parameters regarding the compression was done subjectively. After the data analysis steps mentioned above, the MOS, CI, and JOD values were obtained, as presented in Fig. 5. For better visualisation, different input point counts are shown with different line styles. As can be seen from the plots, there is a small difference between MOS and JOD values, especially for the videos with higher input point count. This is due to the presence of the uncompressed reference in the DSIS test. The degradation was more visible in these cases, whereas there was no reference in the PWC experiment. The relationship between MOS and JOD values is almost linear as shown in Fig. 5.e, and both experiments validate each other. It can be seen that there is a logical distribution of subjective quality values with varying bitrate.

To analyse the significance of the differences between different input point counts, a binomial test was conducted. The results of this binomial were aggregated for each input point count pair for the videos with the same QPs and different input point counts and reported in Fig. 6. In this figure, considering 8 pairs of 127K-

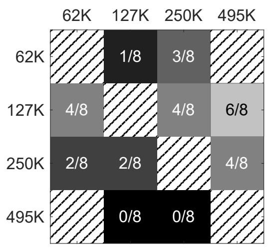


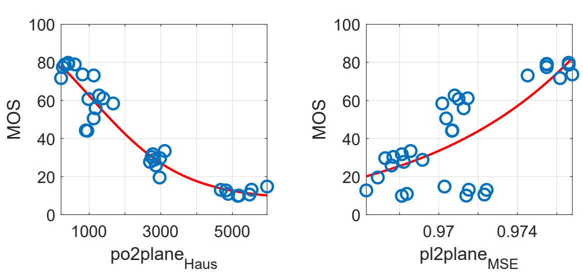
Figure 6. *Binomial test results for different input point counts. Each cell indicates the ratio of the videos with point counts given in the i^{th} row is significantly better than j^{th} column. For this analysis, only the comparisons between videos with the same QPs and different input point counts were considered. These comparisons were aggregated together for each point count pair (e.g. 127K-250K), which makes a total of 8 cases (4 QPs $\times$ 2 contents) for each cell of the table. Striped cells were not compared.*

250K, we see that in two cases 250K was better and in four cases 127K was better. In the remaining two cases, there was not a significant result. Although it is not clear why the videos with 127K are found to be significantly better than other (especially 495K) point counts, we can say that there is no direct correlation between the input point counts and the users' preferences.

As the qualitative part of these experiments, the subjects were asked of their opinions at the end of the experiments, in an informal way. Most of them mentioned that faces were very important for them, which is in agreement with the previous findings [11]. Also, most of the subjects preferred 'Rafa' since he moved less compared to 'Matis' and thus had fewer acquisition errors.

Evaluation of objective quality metrics

Since the DSIS experiment is closer to a 'full-reference' assessment and the relationship between MOS and JOD is almost linear (see Fig. 5.e), we use MOS values in this subsection for the evaluation of objective metrics. The quality scores estimated by the objective metrics were plotted against the MOS values. To evaluate the objective quality estimates, a non-linear logistics function [29] –i.e. $MOS_{pred} = \beta_2 + (\beta_1 - \beta_2)/(1+e^{-(Q_{obj}-\beta_3)/|\beta_4|})$– was used and the Pearson correlation coefficient (PCC), Spearman rank-ordered correlation coefficient (SROCC), root mean square error (RMSE), and outlier ratio (OR) were computed [28]. Sample plots are shown in Fig. 7 and corre-

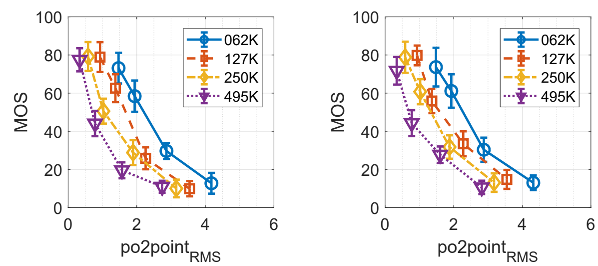


(a) $po2plane_{Haus}$ vs. MOS (b) $pl2plane_{MSE}$ vs. MOS

Figure 7. *Plots for the the best-performing (a) and the least-performing (b) point-based geometry-difference objective quality assessment results. The red lines indicate the best-fit for the non-linear mapping function.*

Metrics	**PCC**	**SROCC**	**RMSE**	**OR**
$po2point_{RMS}$	0.847	0.819	13.37	0.031
$po2point_{Haus}$	**0.958**	**0.913**	**7.24**	**0**
$po2plane_{RMS}$	0.944	**0.912**	8.27	**0**
$po2plane_{Haus}$	**0.958**	0.911	**7.17**	**0**
$pl2plane_{Mean}$	0.767	0.629	16.13	0.125
$pl2plane_{RMS}$	0.764	0.634	16.21	0.125
$pl2plane_{MSE}$	0.764	0.629	16.21	0.125
$PSNR\text{-}po2point_{RMS}$	0.923	0.887	9.65	**0**
$PSNR\text{-}po2point_{Haus}$	0.919	0.887	9.92	**0**
$PSNR\text{-}po2plane_{RMS}$	0.931	0.895	9.16	**0**
$PSNR\text{-}po2plane_{Haus}$	0.930	0.888	9.24	**0**
$PSNR_{YUV}$	**0.990**	**0.969**	**3.56**	**0**

Table 1 - The computed correlation coefficients to evaluate the performances of various objective quality metrics. The bold entries show the three best results each column.



(a) 'Matis' (b) 'Rafa'

Figure 8. *The plot of $po2point_{RMS}$ vs. MOS for both of the contents. Despite having a $PCC = 0.847$, the objective quality estimates are misleading and do not correspond MOS values well due to high variance. For instance, for the 1.75 $po2point_{RMS}$ value, MOS values change from 20 to 75, which indicates very high variance.*

lation coefficients are reported in Table 1. We notice that metrics computed using Hausdorff distance are performing better than others.

The results show that the colour metric $PSNR_{YUV}$ is the most correlated objective metric, followed by $po2plane_{Haus}$. Considering the correlation results, we can say that some of the geometric quality metrics are not performing well. The poor performance of $pl2plane$ metric can be due to the estimation of point normals. As we did not have the normals for each point after decompression, they were estimated using Matlab's `pcnormals` function. This function may be performing badly for very distorted PCs, but this can be also considered as a weakness for the $pl2plane$ metric. Another example is $po2point_{RMS}$ metric. Although the correlation results were not bad, the quality estimates for videos with very similar MOS values vary substantially, as shown in Fig. 8. The results may also indicate that the texture is more important for human perception, but we cannot make this claim since the data is very limited.

Conclusion

In this study, we compress two volumetric videos represented as PCs with colour information, using MPEG PCC TMC2 algorithm. The effects of this compression algorithm and input point counts on the perceived quality were analysed through subjective experiments. The state-of-the-art objective PC metrics were also evaluated.

Results show that, although it has a great effect on geometric quality metrics, the input point count does not affect the perceived quality. Hence, there is no clear advantage or disadvantage of using a higher point count as long as the presentation is plausible for human viewers. Some of the objective quality metrics seem to be affected by the strong changes in the PCs due to high compression, and texture is more important –especially for human faces– compared to geometric distortions. It is worth noting that these conclusions are valid for TMC2 compression and the database used. As a future work, this database is going to be expanded and some of the claims of this paper will be further tested.

Acknowledgements

The authors thank Dr. Rafael Pagés, Dr. Konstantinos Amplianitis, Dr. Jan Ondřej, and Dr. Matis Hudon for providing the volumetric videos. This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under the Grant Number 15/RP/2776.

References

[1] D. S. Alexiadis, D. Zarpalas, and P. Daras, "Real-time, full 3-D reconstruction of moving foreground objects from multiple consumer depth cameras," *IEEE Trans. on Mult.*, vol. 15, no. 2, pp. 339–358, Feb 2013.

[2] A. Collet, M. Chuang, P. Sweeney, D. Gillett *et al.*, "High-quality streamable free-viewpoint video," *ACM Trans. on Graph.*, vol. 34, no. 4, pp. 69:1–69:13, Jul. 2015.

[3] R. Pagés, K. Amplianitis, D. Monaghan, J. Ondřej, and A. Smolić, "Affordable content creation for free-viewpoint video and VR/AR applications," *J. of Vis. Comm. and Image Repr.*, vol. 53, pp. 192–201, 2018.

[4] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Trans. on CSVT*, vol. 27, no. 4, pp. 828–842, 2016.

[5] N. O'Dwyer, N. Johnson, R. Pagés, J. Ondřej, K. Amplianitis, E. Bates, D. Monaghan, and A. Smolić, "Beckett in VR: Exploring narrative using free viewpoint video," in *ACM SIGGRAPH 2018 Posters*, ser. SIGGRAPH '18. ACM, 2018.

[6] E. Alexiou, E. Upenik, and T. Ebrahimi, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *MMSP*. IEEE, Oct 2017.

[7] J. Zhang, W. Huang, X. Zhu, and J.-N. Hwang, "A subjective quality evaluation for 3D point cloud models," in *Int. Conf. on Audio, Language and Image Proc.* IEEE, Jul 2014, pp. 827–831.

[8] E. Alexiou and T. Ebrahimi, "On subjective and objective quality evaluation of point cloud geometry," in *QoMEX*. IEEE, Jun 2017.

[9] ——, "On the performance of metrics to predict quality in point cloud representations," in *Applications of Digital Image Processing XL*. SPIE, 2017.

[10] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of compressed point clouds," in *MMSP*. IEEE, Oct 2017.

[11] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," in *Applications of Digital Image Processing XLI*. SPIE, 2018.

[12] E. Alexiou, M. V. Bernardo, L. A. da Silva Cruz *et al.*, "Point cloud subjective evaluation methodology based on 2D rendering," in *QoMEX*. IEEE, 2018.

[13] E. Alexious, A. M. Pinheiro, C. Duarte *et al.*, "Point cloud subjective evaluation methodology based on reconstructed surfaces," in *Applications of Digital Image Processing XLI*. SPIE, 2018.

[14] E. Alexiou and T. Ebrahimi, "Impact of visualization strategy for subjective quality assessment of point clouds," in *ICME Workshops*. IEEE, 2018.

[15] R. Mekuria, Z. Li, C. Tulvan, and P. Chou, "Evaluation criteria for PCC (Point Cloud Compression)," ISO/IEC JTC 1/SC29/WG11 Doc. N16332, 2016.

[16] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *ICIP*. IEEE, Sept 2017, pp. 3460–3464.

[17] E. Alexiou and T. Ebrahimi, "Point cloud quality assessment metric based on angular similarity," in *ICME*. IEEE, 2018.

[18] K. Mammou, "PCC Test Model Category 2 v0," ISO/IEC JTC 1/SC29/WG11 Doc. N17248, 2017.

[19] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in *Int. Conf. on Rob. and Aut. (ICRA)*. IEEE, May 2011.

[20] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. on CSVT*, vol. 22, no. 12, pp. 1649–1668, 2012.

[21] ITU-T, "Subjective video quality assessment methods for multimedia applications," ITU-T Rec. P.910, Apr 2008.

[22] M. Perez-Ortiz and R. K. Mantiuk, "A practical guide and software for analysing pairwise comparison experiments," 2017.

[23] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," ITU-R Rec. BT.500-13, Jan 2012.

[24] M. Kleiner, D. Brainard, D. Pelli, A. Ingling *et al.*, "Whats new in Psychtoolbox-3," *Perception*, vol. 36, no. S, 2007.

[25] E. Zerman, V. Hulusic, G. Valenzise, R. K. Mantiuk, and F. Dufaux, "The relation between MOS and pairwise comparisons and the importance of cross-content comparisons," in *IS&T Electronic Imaging, HVEI XXII*, Jan 2018.

[26] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards – including high efficiency video coding (HEVC)," *IEEE Trans. on CSVT*, vol. 22, no. 12, pp. 1669–1684, Dec 2012.

[27] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Evaluation metrics for point cloud compression," ISO/IEC JTC 1/SC29/WG11 Doc. M39966, 2017.

[28] ITU-T, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," ITU-T Rec. P.1401, Jul 2012.

[29] A. M. Rohaly, J. Libert, P. Corriveau, A. Webster *et al.*, "VQEG final report of FR-TV phase I validation test," http://www.vqeg.org/, Mar 2000.

Author Biography

Emin Zerman is a postdoctoral research fellow at Trinity College Dublin, working in V-SENSE project. He received his B.Sc. degree (2011) and M.Sc. degree (2013) in Electrical and Electronics Engineering from the Middle East Technical University, Turkey, and his Ph.D. degree (2018) in Signals and Images from Télécom ParisTech, France. His research interests include high dynamic range imaging, immersive multimedia, human visual perception, computer vision, video processing, and multimedia quality assessment.