2021 Thirteenth International Conference on Quality of Multimedia Experience (QoMEX)

A Full- and No-Reference Metrics Accuracy Analysis for Volumetric Media Streaming

Sam Van Damme, Maria Torres Vega and Filip De Turck
IDLab, Department of Information Technology (INTEC)
Ghent University - imec
sam.vandamme@ugent.be

2021 13th International Conference on Quality of Multimedia Experience (QoMEX) | 978-1-6654-3589-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/QoMEX51781.2021.9465420

***Abstract*—Volumetric media streaming will be one of the fundamental technologies to enable near future immersive multimedia experiences. In it, objects represented as sets of points (*i.e.* point-clouds), are presented to remote users wearing Head-Mounted Displays (HMDs). Due to the stringent bandwidth and latency requirements of such applications, small changes in the network can affect the user in unexpected manners (physical discomfort, lack of concentration, etc.). Therefore, there is a need for assessing the perceived quality of this type of applications in real-time, i.e, the Quality of Experience (QoE). Given that subjective evaluations are not feasible for (near) real-time applications, objectively measuring this quality will be a must. While traditional objective metrics could potentially be used to fulfill the task, it is still unclear how accurate they are to assess volumetric media. To this end, this paper presents a thorough correlation analysis of both Full Reference (FR) and No Reference (NR) objective metrics to subjective Mean Opinion Scores (MOS) for different volumetric streaming scenarios. To enhance the accuracy, multiple Region-Of-Interest (ROI) selection and weighting procedures have been applied and their influence on the results have been investigated. Our results show that the classical video quality metric Video Multimethod Assessment Fusion (VMAF) is well-suited as an objective benchmark for volumetric media streaming in terms of correlation to subjective scores, while a combination of NR features could provide a suitable real-time assessment. Finally, ROI selection proves to widen the range of objective metrics, which is an important issue to apply traditional objective metrics to volumetric media.**

***Index Terms*—Volumetric media, Quality of Experience, No Reference, Full Reference, objective quality, Region-Of-Interest**

I. Introduction

The significant increase on popularity of Augmented and Virtual Reality (AR/VR) content and applications has made network and content providers to start offering their content with 3 or 6 Degrees of Freedom (DoF), where the user is immersed in a virtual environment which they can explore and interact with. One clear example of such applications is point cloud delivery [1]. In these, objects composed by a dense network of 6D points (x, y, z + three colour channels) are presented to the user's Head-Mounted Display (HMD). The users can move around and interact with the figures, and explore them from different sides and angles. However, stringent requirements on the network, *e.g.* bandwidth, can result in low quality rendering, *i.e.* a reduced number of points in the cloud, blurriness or freezes. These effects can highly affect the user's level of immersiveness and degrade their perception of the application. Therefore, there is a need for continuous and real-time monitoring of the Quality of Experience (QoE) to manage these applications.

Objective metrics would be best suited for this analysis, as no human intervention in the form of subjective experiments is needed. To this end, it is highly beneficial to have an objective Full Reference (FR) benchmark that correlates well to subjective Mean Opinion Scores (MOS). Moreover, it is possible to create No Reference (NR) quality models to perform low-complexity quality estimation on light-weight client devices, using the benchmark for training purposes. However, there is currently no clear consensus in literature on which benchmarks and NR features to use for these particular purposes. In addition, research towards the accuracy and applicability of traditional NR video features for quality modelling is non-existent. Besides, the existing metrics include the background in their calculations [2]–[4], while one can expect that the user will primarily focus on the point cloud objects themselves. As such, Region-Of-Interest (ROI) extraction mechanisms are needed in order to adapt existing metrics to be more tailored towards human perception [5].

Therefore, the contribution of this paper is two-fold. First, it focuses on the selection of an objective FR metric that correlates well with subjective results such that it can act as an alternative ground-truth benchmark, in case subjective studies cannot be run. Second, it provides insight into the applicability and accuracy of pixel-based NR features for real-time quality assessment of volumetric media, taking into account the benefits of suitable background extraction.

The remainder of this paper is distributed as follows. Section II provides a brief description of the background on FR, Reduced Reference (RR) and NR metrics applied to volumetric media delivery. Section III presents an overview of the experimental methodology followed for the evaluation. Section IV discusses the subjectively annotated dataset used as well as the results. Finally Section V concludes the paper.

II. Background

Several objective metrics have been used in literature to assess the quality of a volumetric object. These can be divided in two major classes, *i.e.* the quality of the point cloud object itself (*geometric*) and the quality of the rendered Field-of-View (FoV), *i.e.* the content the user observes when looking

around using a HMD (*projection-based*). The former are often calculated as point-to-point or point-to-plane Peak-Signal-to-Noise-Ratio (PSNR)-based distortion metrics [6]. Although these give insight about the performance of the applied compression, they do not provide an indication of the user's visual quality perception [7]. To this end, traditional video quality metrics have recently been investigated to assess the quality of the user's FoV. In terms of accuracy, the FR metrics (which provide a full comparison between the original and distorted sequence) show the highest potential. These include PSNR, Structural Similarity Index Measure (SSIM) [2], Video Quality Metric (VQM) [8] and Netflix' Video Multimethod Assessment Fusion (VMAF) [1], to name a few. Due to the computationally higher complexity of most FR metrics and because simultaneous access of both the original and distorted content is needed, FR metrics are not applicable to end-user real-time evaluations. These problems can be potentially bypassed by applying NR metrics as a measurement of quality, which make an assessment purely on the distorted stream. Note that each of these metrics take the background of the consumed scenes into account, which is assumed to contribute less to the perceived quality, since it is expected that users mainly focus on the objects in the front, *i.e.* the ROIs. One work has considered background removal for images generated from point cloud content, using a MATLAB-based tool for assisted removal that thresholds on the transparency values in RGBA space [5].

In recent years, multiple attempts have been made to tailor these *projection-based* traditional metrics to the specific characteristics of volumetric media. Yang et al. [9] presented a FR metric by first projecting the 3D point cloud on the six perpendicular planes of a cube. Next, for each of the six resulting images, features are extracted from both the colour information and the depth map in terms of edges, texture similarity, and Jensen-Shannon (JS) divergence. Afterwards, the six contributions are weighted to obtain one overarching quality index. Their results show Pearson Linear Correlation Coefficients (PLCCs) with subjective MOS ranging from 0.66 to 0.97 depending on the particular content and the encoding distortions introduced.

Diniz et al. [10] derived a RR point cloud quality assessment model based on local binary patterns. To this end, a binary code is associated to each pixel by thresholding the difference in intensity with its surrounding pixels. The quality of the point cloud is then calculated based on the difference between the histograms of the original and distorted content by mapping this distance to a predicted MOS using a 3rd order polynomial relationship. Their results show PLCCs to MOS varying between 0.667 and 0.876, depending on the database.

Viola et al. [11] created a RR color-based quality metric by extracting colour statistics in terms of histograms and correlograms from both the reference and the degraded content. The distance between both is used as a predictor for subjective MOS by applying a curve-fitting approach. Their results show a PLCC up to 0.904 with MOS by applying a logistic curve-fitting. In a second study [12], they present another RR metric based on a weighted combination of feature differences in terms of geometry, luminance and normal. Evaluation on a publicly available dataset shows PLCCs between 0.798 and 0.901 to subjective MOS, depending on the followed Cross-Validation (CV) approach.

Alexiou et al. [13] adapted the traditional SSIM for use in point clouds based on geometry, normal vectors, curvature values and colors. They reach maximum PLCCs to MOS between 0.8 and 0.9, depending on the dataset.

In our own, previous work [14], at last, we presented an objective and subjective quality evaluation of point cloud streaming for multiple scenarios in terms of bandwidth, rate adaptation, viewport prediction and user motion. The results show high correlation with MOS for traditional video metrics such as PSNR, SSIM and VQM. We further indicated that the subjective perception of volumetric media lays within a very small interval of the total range of the objective metrics, which might be a result of the inclusion of (too much) background during the quality metric calculation.

As can be noticed from this discussion, there is no real consensus on which FR metrics to use as a benchmark for projection-based approaches, as all of them are directly comparing to subjective scores. This is highly unpractical, though, as it would require a costly subjective study for every new volumetric media sequence being added to the database. Moreover, research towards the accuracy and applicability of traditional NR features to volumetric media delivery is non-existent. Finally, additional research is needed on background removal and/or ROI selection and their impact on quality metric correlations and range. As such, objective metrics can be guided towards a more accurate representation of subjective human perception.

III. Experimental Methodology

The purpose of this work is two-fold: (i) to find a good FR benchmark, well correlated to subjective evaluations and that could potentially be used as ground-truth benchmark if there were no subjective results available; (ii) to understand if pixel-based NR features can provide a real-time assessment of quality degradation for volumetric media. To achieve this double objective, the presented approach was followed (Figure 1). A server stores a set of volumetric (point-cloud) objects in different quality variants. The figures are integrated into a video, which is streamed over an emulated network. This network can provide different bandwidths and latency constraints under controlled conditions. The video stream, impaired by the network, is received at the end-user device, where subjective evaluations are driven. These can be done both as Double Stimulus (DS) (side-by-side comparison of original unimpaired and impaired streams) as well as Single Stimulus (SS) (where the subjects rate the presented stream in one single screen). At the same time, FR and NR quality evaluations are performed. To do so, the background is extracted and the volumetric figures are analysed on the pixel-level. The objective features are then benchmarked against the subjective scores. Three types of

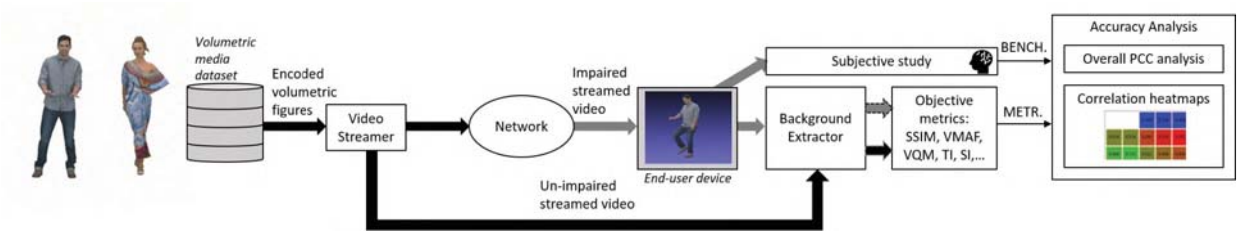


Fig. 1: Experimental methodology block diagram.

TABLE I: Objective features used for this experimental evaluation. For each of them, type (FR or NR), name, acronym and a brief description are provided.

Type	Acronym	Name	Description
FR	VMAF	Video Multimethod Assessment Fusion	Netflix's FR quality metric. It is forged out of four features (ANSNR, DLM, VIF and MI) using a Support Vector Regression (SVR) approach with subjective MOS as the benchmark [1].
	SSIM	Structural Similarity Index Measure	Models the perceived change in structural information based on the strong interdependencies of spatially close pixels. The luminance, colour and structure of the frame are taken into account to this end [2].
NR	BLU	Average blur	Blurred pixels are identified by thresholding on the difference between a pixel and the corresponding pixel in the derivative image. The average blur is this average difference divided by the total amount of blurred pixels. The per-frame values are averaged over the video sequence afterwards [3].
	BRT	Blur ratio	Ratio of the number of blurred pixels (derived as in BLU) to the total amount of edge pixels, after applying an edge detection algorithm. The per-frame values are averaged over the video sequence afterwards [3].
	NOI	Average noise	Noisy pixels are identified by thresholding on the difference between the local derivative and the average derivative. The average noise is this average difference divided by the total amount of noisy pixels. The per-frame values are averaged over the video sequence afterwards [3]
	NRT	Noise ratio	Ratio of the number of noisy pixels (derived as in NOI) to the total number of pixels in the image. The per-frame values are averaged over the video sequence afterwards [3].
	BLO	Blockiness	Calculated by analysing the inner and outer edges of 8x8 subblocks on both the vertical and horizontal Sobel-filtered versions of the frame. As such, an *inner* and *edge blockiness* level is determined, of which the average difference over all blocks describes the blockiness value of the frame. The per-frame values are averaged over the video sequence afterwards [4].
	SI	Spatial Information	Measurement for the degree of spatial detail, calculated by taking the standard deviation of the pixel intensities of a Sobel-filtered version of each frame. Next, the maximum of this set of standard deviations is taken to represent the Spatial Information (SI) of the sequence [15].

analysis are envisioned: overall PLCC correlations (to get an idea of the general applicability of the metric, and the linearity between the metric and the benchmark), curve fitting (due to the often sigmoidal relationship between objective quality and subjective perception [16]), and correlation colour maps (to compare the per-video performance for multiple features and benchmarks). The remainder of this section provides a description of the background extraction procedure selected, as well as of the FR and NR features employed for the analysis.

A. *Background extractor*

Background extraction is realised using the statistical estimation and per-pixel Bayesian segmentation algorithms as proposed by Godbehere et al. [17]. The resulting ROIs are cut from the frame by identifying the smallest enveloping rectangle. These background masks only need to be calculated on the reference videos, as the resulting binary maps can straightforwardly be applied to the distorted sequences covering the same content as the reference video. Based on the resulting ROIs, multiple approaches are put forward for metric calculation.

- *STA:* The standard approach. No background extraction or ROI selection is performed and quality metrics are calculated on the full frames.
- *AVG:* Metrics are calculated for each ROI separately. Afterwards, the quality indexes are averaged (based on the assumption that users pay equal importance to each of the ROIs) to obtain a quality score for the whole frame.
- *CEN:* Metrics are calculated on the ROI closest to the center of the frame, assuming that the user will only focus on this silhouette. Other ROIs are neglected.
- *CENG:* Metrics are calculated for each ROI separately. Afterwards, they are weighted using a Gaussian weight function over the horizontal axis of the frame with its mean at the center of the ROI closest to the center of the frame. The standard deviation is set to 133,3 such that the full width of the frame covers 99.7% of the area under the curve. This is based on the assumption that users will pay more attention to the center of their gaze than on the edges, but that the latter cannot be neglected completely.
- *NEW:* Metrics are calculated on the ROI that appeared most recently in the video, assuming that the user will always shift his/her focus once a new silhouette appears. Other ROIs are neglected.
- *NEWG:* Metrics are calculated for each ROI separately. Afterwards they are weighted using a Gaussian weight function similar to the *CENG* approach. This time, however, the mean of the Gaussian is placed at the center of the ROI that most recently appeared.

B. Objective metrics

For the accuracy analysis we focused on two well-known FR metrics, namely VMAF and SSIM, as well as five pixel-based NR features. Table I provides a summary of the different metrics under scrutiny. First, VMAF was selected due to its good correlation to MOS for 2D videos [18]. As it works on the video level, it cannot be used to provide a frame-by-frame analysis, which would be required for real-time assessment of quality. However, it could provide a valuable benchmark as an alternative to subjective evaluations. SSIM, on the other hand, provides a per-frame analysis of the structures within the frame. This could provide a decent frame-by-frame analysis, but the state of the art has shown that SSIM's working range is rather low, due to the influence of the background on each of the frames [14]. Background removal procedures could improve its working range as well as its correlation to subjective evaluations.

NR features, on the other hand, make an assessment purely on the distorted stream. In this paper, the focus was set on low complexity pixel-based features, which can be run on light-weight devices in real-time as new frames arrive. Among all possible features, we selected noise (average and ratio), blockiness, blur (average and ratio) and the SI of the frames. Each of them were implemented in Python following state of the art implementation, similarly as in our previous work [16]. Furthermore, as the first five provide a measurement of degradation, instead of quality, they were inversed and set between 0 and 1, where 0 means full degradation and 1, full quality. In that case, they could be easily correlated to the subjective scores.

IV. Evaluation

This section presents the dataset used for the evaluation. Then it shows the results and answers the research questions.

A. Dataset & subjective evaluations

We used the subjectively annotated dataset as created in our previous work [7]. It evaluates the subjective perception of users when volumetric media is streamed under changing adaptive conditions. Users were shown a number of source videos between 18 and 24 seconds of length, containing the generated viewport of a scene consisting of four point cloud objects from the 8i dataset [19]. These objects were encoded using the V-PCC encoder with five reference quality representations, each between 2.4 Mb/s and 53.5 Mb/s. For these evaluations, three types of video scenes were considered, each with a different setup of the figures (line vs. semicircle) and camera movement (panning vs. zoom-in-zoom-out). A total of eight configurations per video were selected resulting in a total number of 24 test video sequences which were subjectively assessed by each user in random order on a 2D screen. A total of 60 subjects participated in two subjective experiments, namely 30 for the SS and the other 30 for the DS. For the results shown in the next Section, both analyses were used.

TABLE II: PLCC correlations for VMAF and SSIM per video and overall for the two flavors of subjective evaluations.

	MOS_SS		MOS_DS	
Video	VMAF	SSIM	VMAF	SSIM
Video 1	0.7	0.67	0.84	0.81
Video 2	0.95	0.92	0.97	0.94
Video 3	0.94	0.91	0.99	0.98
Total	0.88	0.8	0.93	0.85

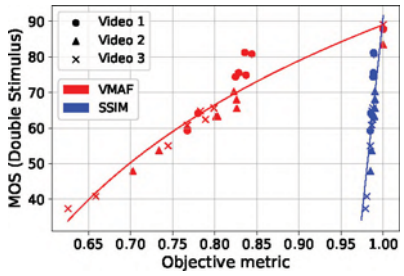
B. Results

In the next subsections, the two research questions are answered.

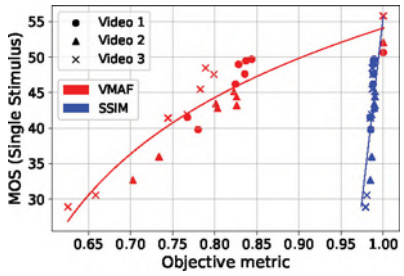
1) Can FR metrics be used as alternative when subjective evaluations are not available?: To answer this question, a PLCC evaluation was performed of both VMAF and SSIM, for the three video types and the two flavors of subjective metrics. As is shown in Table II, both SSIM and VMAF provide high correlation (between 0.7 and 1) for both the SS and DS evaluations. As was to be expected, the correlations are higher for the DS evaluations. This is because a DS subjective evaluation is closer to a FR assessment, i.e., the user is allowed to compare the unimpaired and impaired content. Another interesting conclusion is that the PLCCs are relatively worse for video 1. This could be due to the type of structure of the video. While video 1 provides a panoramic view of the four figures, videos 2 and 3 consist of a zoom in and out of two out of the four. Thus, it seems that the structure of video 1 makes it more difficult for the FR metrics to assess quality in line with the user's perception. Enhancements on the metrics to follow this type of video content would be useful to increase the accuracy, but in general, it can be concluded that VMAF provides a good benchmark even though as a video based approach it cannot be used at a frame-level.

The next step was to understand if the FR could be used for the evaluation in terms of the working range. Having a very short working range would make the metric not to provide an insightful value of quality. Figure 2 presents the curve fitting results for the two FR metrics against the MOS values for both DS (Figure 2a) and SS (Figure 2b). As can be seen, the VMAF evaluations show a range of 0.35 between the lowest and highest quality, while SSIM's barely reaches 0.03. This makes VMAF very well suited for objective evaluations as a large scale alternative to subjective evaluations. SSIM would be better suited for online evaluations (as the frames are being received). Therefore, we applied the different background extractor algorithms to try to increase the working range. Figure 3 presents the curve fitting of SSIM against DS and SS MOS. In these figures, the video markers have been removed for clarity. As can be seen, removing the background of the videos clearly increases the working range of SSIM, where the *NEWG* approach improves the range of SSIM from the original 0.03 to 0.15, while the overall PLCC (Table III) does barely suffer. This shows the potential of background extraction to adapt traditional FR metrics to volumetric projected media streaming.

2) Can traditional NR features be used for real-time assessment?: As a second step, we aimed to understand if

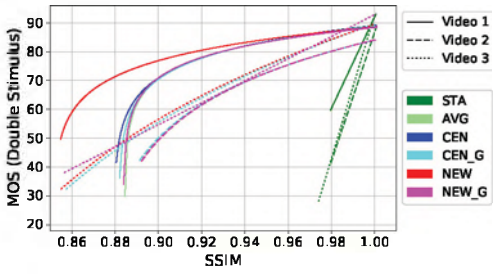


(a) DS MOS

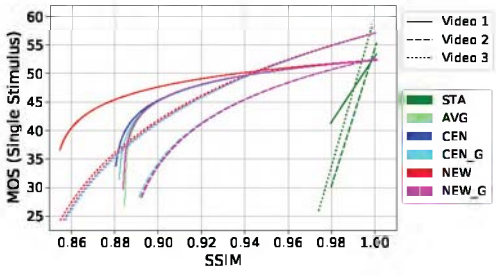


(b) SS MOS

Fig. 2: Curve fitting of the values provided by VMAF and SSIM to the subjective evaluations of the provided dataset.



(a) DS MOS



(b) SS MOS

Fig. 3: Curve fitting of the SSIM values to the subjective evaluations of the provided dataset, using different background extraction approaches.

traditional NR features could be used for real-time client-based assessment at the pixel-level. As shown in Table III, the overall PLCCs for the selected NR features are rather limited for each of the MOS flavors. We added VMAF as a third benchmark, given its good performance in the previous analysis. There can be seen that overall PLCCs of NR are rather low, with values varying from -0.3 to 0.62. However, applying the proposed ROI selection methods increases or at least levels the strength of the correlation in comparison to *STA* for most of the cases. For the *BRT* feature, for instance, the *CEN* and *CENG* consistently show to improve the correlation compared to the *STA* case for each of the benchmarks. For the *NOI* feature, the same conclusion can be drawn for the *NEW* approach.

As previously pinpointed for SSIM and VMAF, objective metrics often show different behaviour depending on the video at hand. To this end, a per-video correlation colour map analysis to each of the three benchmarks is performed (Figure 4). First of all, strong PLCCs can be noticed of *BRT* and *NOI* to each of the benchmarks and for each of the three videos, with values up to 0.97 and 0.96, respectively, depending on the video and the ROI selection approach. In addition, it should be emphasised that the obtained PLCCs of these two features are mostly increasing when compared to *STA*, proving the added value of ROI selection. Furthermore, the strong negative and positive correlations of *NRT* and *SI* for video 2 should be noted as well. Especially in case of *NRT*, there is once again a clear improvement by applying ROI selection prior to the calculation of the features. This conclusion is less pronounced for *SI*. A similar conclusion can be made for the strong negative correlations of the *BLU* feature in video 3. This is an interesting result, as the same feature is showing limited positive correlations for video 2 and strong negative to zero approximate correlation values for video 1.

TABLE III: Overall PLCC correlations of SSIM and the NR features to the SS MOS, DS and VMAF evaluations. The SSIM values are shown in blue, while the best and worst correlation values (in absolute value) of the NR features in all three analyses are shown in bold green and italic red respectively.

	App.	SSIM	BLU	BRT	NOI	NRT	BLO	SI
MOS SS	STA	0.8	-0.23	0.57	0.39	0.25	0.19	0.3
	AVG	0.77	-0.3	0.56	0.39	0.25	-0.05	0.1
	CEN	0.76	-0.25	**0.59**	0.37	0.25	0.13	0.23
	CENG	0.77	-0.29	0.59	0.32	0.25	0.12	0.22
	NEW	0.67	-0.28	0.56	0.4	-0.28	0.14	0.3
	NEWG	0.77	-0.3	0.56	0.35	0.23	*0.01*	0.13
MOS DS	STA	0.85	-0.42	0.59	0.40	0.45	0.12	0.17
	AVG	0.81	-0.48	0.60	0.39	0.45	0.01	-0.08
	CEN	0.8	-0.46	0.62	0.34	0.44	0.29	0.13
	CENG	0.80	-0.51	**0.62**	0.26	0.43	0.28	0.11
	NEW	0.68	-0.48	0.61	0.44	-0.27	0.29	0.26
	NEWG	0.8	-0.48	0.60	0.31	0.43	0.10	*-0.04*
MOS DS	STA	0.97	-0.39	0.38	0.21	0.27	-0.01	0.07
	AVG	0.95	-0.29	0.39	0.20	0.26	0.11	-0.11
	CEN	0.95	-0.36	0.42	0.16	0.25	0.26	0.01
	CENG	0.95	-0.37	**0.42**	0.10	0.24	0.24	*-0.00*
	NEW	0.86	-0.31	0.40	0.23	-0.09	0.27	0.09
	NEWG	0.95	-0.29	0.40	0.14	0.25	0.16	-0.09

To summarise, it can be concluded that the presented ROI selection procedures surely show their potential to improve on traditional NR metrics in the context of volumetric media. In addition, the correlations show that an accurate and lightweight NR quality model for point clouds is within reach. On the downside, the best suited NR features and ROI selection approach tend to vary over the multiple videos rather than

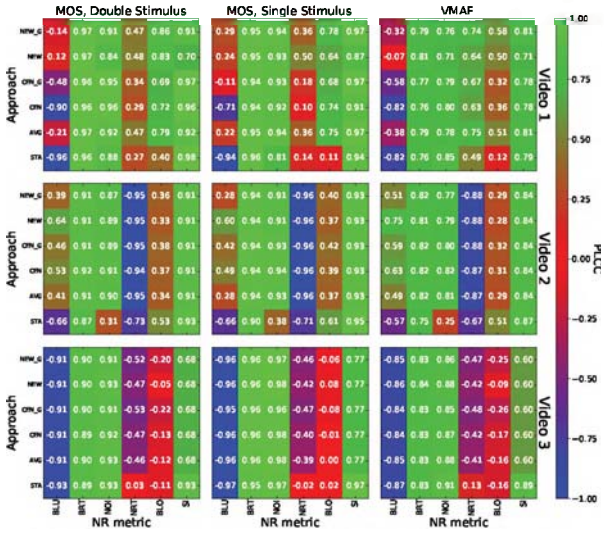


Fig. 4: PLCC correlation colour maps of the NR features to the three benchmarks, namely the MOS SS, the MOS DS and VMAF, using different background extraction approaches.

showing a general optimum. How to determine these is therefore an open issue of research. A possible solution would be a combination of NR features and content dependent background extraction procedures. Furthermore, additional research should be performed on the combination of multiple NR features into one, accurate metric. Machine Learning (ML) approaches might provide a useful tool to this end.

V. Conclusion

Volumetric media streaming will be one of the core applications of near future immersive multimedia experiences. Providing real-time assessments of the perception of this type of service is still an open research question. In this paper we have presented a thorough correlation analysis of both FR and NR objective metrics to subjective MOS with a double purpose: (i) can objective FR metrics be used as an alternative when subjective evaluations are not available? (ii) is it possible to provide real-time accurate assessment of volumetric media with low computation NR features? In addition, to enhance the accuracy, we have explored the effects of ROI selection and weighting procedures on the accuracy results. Our results have shown that the classical video quality metric VMAF is very well-suited as an objective benchmark for volumetric media streaming in terms of correlation to subjective scores. Moreover, a combination of NR features could provide a good real-time assessment. Finally, ROI selection has proven to widen the range of objective metrics, both for FR (SSIM) as well as NR features. This has been pointed out as a fundamentally important issue to apply traditional objective metrics to volumetric media. How to determine the best suited NR features and ROI selection procedure for a given video is still an open issue of research, which we aim to investigate as part of our future work. In addition, research will be performed towards the application of ML solutions for the creation of an accurate NR metric out of the presented features.

Acknowledgements

This research is part of a collaborative project between Huawei and Ghent University, funded by Huawei Technologies, China. Maria Torres Vega is funded by the Research Foundation Flanders (FWO), grant number 12W4819N.

References

[1] A. Aaron, Z. Li, M. Manohara, J. Y. Lin, E. C. Wu, and C. . J. Kuo. Challenges in cloud based ingest and encoding for high quality streaming media. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 1732–1736, 2015.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Processing*, 13(4), 2004.

[3] Min Goo Choi, Jung Hoon Jung, and Jae Wook Jeon. No-reference image quality assessment using blur and noise. *International Journal of Computer Science and Engineering*, 3(2):76–80, 2009.

[4] C. Perra. A low computational complexity blockiness estimation based on spatial analysis. In *2014 22nd Telecommunications Forum Telfor (TELFOR)*, pages 1130–1133, 2014.

[5] E. Alexiou and T. Ebrahimi. Exploiting User Interactivity in Quality Assessment of Point Cloud Imaging. In *QoMEX*, 2019.

[6] MPEG. MPEG 3DG and Requirements - Call for Proposals for Point Cloud Compression V2. https://bit.ly/2RSWdWe, 2017.

[7] J. van der Hooft, M.Torres Vega, T. Wauters, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz. From capturing to rendering: Volumetric media delivery with six degrees of freedom. *IEEE Communications Magazine*, 58(10):49–55, 2020.

[8] M. H. Pinson and S. Wolf. A New Standardized Method for Objectively Measuring Video Quality. *IEEE Transactions on Broadcasting*, 50(3):312–322, 2004.

[9] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang, and J. Sun. Predicting the perceptual quality of point cloud: A 3d-to-2d projection-based exploration. *IEEE Transactions on Multimedia*, pages 1–1, 2020.

[10] R. Diniz, P. G. Freitas, and 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)*, pages 1–6, 2020.

[11] I. Viola, S. Subramanyam, and P. Cesar. A color-based objective quality metric for point cloud contents. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, 2020.

[12] I. Viola and P. Cesar. A reduced reference metric for visual quality evaluation of point cloud contents. *IEEE Signal Processing Letters*, 27:1660–1664, 2020.

[13] E. Alexiou and T. Ebrahimi. Towards a point cloud structural similarity metric. In *2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, pages 1–6, 2020.

[14] J. van der Hooft, M. Torres Vega, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz. Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, 2020.

[15] ITU-T Rec. P.910: Subjective video quality assessment methods for multimedia applications, 2008.

[16] Sam Van Damme, Maria Torres Vega, Joris Heyse, Femke De Backere, and Filip De Turck. A low-complexity psychometric curve-fitting approach for the objective quality assessment of streamed game videos. *Signal Processing: Image Communication*, 88:115954, 2020.

[17] A. B. Godbehere, A. Matsukawa, and K. Goldberg. Visual tracking of human visitors under variable-lighting conditions for a responsive audio art installation. In *2012 American Control Conference (ACC)*, pages 4305–4312, 2012.

[18] R. R. Ramachandra Rao, S. Göring, W. Robitza, B. Feiten, and A. Raake. Avt-vqdb-uhd-1: A large scale video quality database for uhd-1. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 17–177, 2019.

[19] E. d'Eon, T. Myers, B. Harrison, and P. A. Chou. Joint MPEG/JPEG Input. 8i Voxelized Full Bodies - A Voxelized Point Cloud Dataset. https://jpeg.org/plenodb/pc/8ilabs/, 2017.