The Effect of Temporal Sub-sampling on the Accuracy of Volumetric Video Quality Assessment

Ali Ak, Emin Zerman, Suiyi Ling, Patrick Le Callet, Aljosa Smolic

The Effect of Temporal Sub-sampling on the Accuracy of Volumetric Video Quality Assessment

Ali Ak*, Emin Zerman†, Suiyi Ling*, Patrick Le Callet*, and Aljosa Smolic†
*Image Perception Interaction Team, LS2N, University of Nantes, Nantes, France
Emails: {ali.ak, suiyi.ling, patrick.lecallet}@univ-nantes.fr
†V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
Emails: {emin.zerman, smolica}@scss.tcd.ie

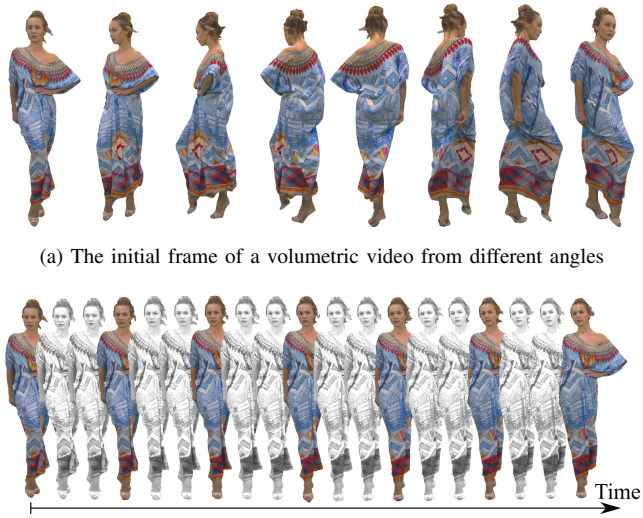
***Abstract*—Volumetric video content has attracted increasing research interests over the last decade, as it facilitates the integration of dynamic real world content in virtual environments. Point cloud is one of the most common alternatives to represent volumetric video content. Yet, such representation requires an enormous data storage and pose significant greater pressures on compression algorithms compared to the standard 2D video. This challenge has unleashed a new wave in the development of novel point cloud compression technologies, which need to be evaluated in terms of production quality. Due to the high dimensionality of the data, evaluating the performances of relevant coding algorithms can be time consuming. This puts a barrier on optimizing coding algorithms with complex, but perceptually accurate, objective quality metrics. In this study, we thus explore the possibility of reducing temporal-dimension of the content under-evaluation, *i.e., temporal sub-sampling*, for objective quality evaluation without sacrificing from the correlation with the subjective opinion. In addition, we exploit different temporal pooling methods to further make the quality evaluation procedure more efficient. In total 30 different objective quality metrics were tested on the the V-SENSE volumetric video quality database. According to experimental results, there is no need to employ full frame-rate (30 fps) assessment to reach the meaningful correlation for the considered quality metrics. These observations could be referred to reduce the computation complexity regarding the evaluation and optimization of the relevant compression algorithms.**

***Index Terms*—temporal sub-sampling, volumetric video, quality assessment, point cloud**

I. Introduction

Recent advancements in acquisition and display technologies enabled a more immersive form of imaging. Being one of these immersive imaging technologies, volumetric video (also known as 3D video or free-viewpoint video) makes it possible to capture 3D objects in the real world and to display them from any angle the viewer wishes to look [1]–[3], see Fig. 1.a. The sequence of reconstructed 3D objects can be stored as textured polygonal meshes or point clouds [4], which can be used in augmented reality and virtual reality scenarios.

Using state-of-the-art techniques, the generated volumetric video sequences can be compressed and transmitted over the Internet. In this context, quality assessment is crucial to ensure that the compression and transmission operations do not degrade the sequence too much or to guarantee highest quality of experience (QoE). Since "quality" is subjective by definition [5], conducting user studies is the golden standard for visual quality assessment. Hence, user studies have been used for subjective quality assessment of volumetric video [4], [6], [7]. Nevertheless, these user studies are both time and resource expensive to conduct.



(a) The initial frame of a volumetric video from different angles

(b) Uniform temporal sub-sampling for 10 fps instead of 30 fps.

Fig. 1. Visualization of (a) a sample volumetric video frame and (b) a uniform temporal sub-sampling scheme, where the sampled frames are shown in color while others are shown in black and white.

Objective visual quality assessment metrics are commonly used to estimate the perceived visual quality without conducting a subjective user study. These metrics generally work for a specific data representation, *e.g.*, image-based metrics work on pixels and point-based metrics work on points in 3D space. There is a plethora of metrics that were designed for traditional images [8], and there are some metrics that were developed for point clouds (Please refer to Section III-B for more details) for volumetric video quality assessment [9]–[11]. Recent studies also show that the volumetric videos can be rendered as images and traditional image quality metrics can be used for quality assessment [12]. The current challenge for objective quality assessment for volumetric video is that the sizes of volumetric

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 765911 (RealVision) and from Science Foundation Ireland (SFI) under the Grant Number 15/RP/27760.

Fig. 2. Visualization of the initial frames of vsenseVVDB2 contents.

video sequences are big, and the objective quality evaluation procedure can thus be time-consuming. Therefore, in this paper, we seek answer to the following question: *"Can we speed up the metric computation for volumetric video quality assessment without sacrificing the accuracy?"*

In the literature, temporal sub-sampling methods have been used to reduce the amount of calculations and increase prediction accuracy in video quality domain [13]–[15]. Tu *et al.* conducted a benchmark study on pooling methods for blind video quality assessment [15]. Authors concluded that pooling method choice is content dependent and an ensemble method could improve the results for such ill-posed problems. Seufert *et al.* [14] suggested that the arithmetic pooling works better than sophisticated pooling methods for sequences of length in order of minutes. However, to the best of our knowledge, these approaches have not been investigated for volumetric video content as the volumetric video with subjective quality annotations were not available until recently [6].

In this paper, we study the effects of temporal sub-sampling on the accuracy of quality assessment models for volumetric video sequences, see Fig. 1.b. In particular, we focus on volumetric videos represented as point clouds. Moreover, we also analyze the effect of temporal pooling methods on the quality assessment of volumetric videos. The results provide insight for speeding up the objective quality estimation for volumetric video sequences, which will be useful for the community.

II. Temporal Pooling Methods

Considered temporal pooling methods are described in this section. F_i is the i_{th} frame in the video with an objective quality score of q_i. i ranges from 1 to N, where N indicates the last frame of the video. Finally, quality score of the video is denoted as Q. Table I lists the pooling methods which is used in the experiment. Formulas and selected values for the adjustable parameters are given in the table.

Arithmetic mean is calculated as the mean value of quality scores across frames within a video. Harmonic mean uses a similar definition with negative exponent to have higher impact on frames with lower quality. Minkowski mean is a generalized version of the arithmetic and harmonic mean with an adjustable parameter. VQ-Pooling was proposed as a pooling method considering both spatial and temporal domain [16]. We use temporal component of VQ-Pooling as its adapted version proposed in [15]. Concretely, individual quality scores of all frames are clustered into two groups utilizing the K-means algorithm, which were denoted as G_L and G_H. Afterwards,

TABLE I
Definitions and selected parameters for pooling methods.

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

the overall quality prediction is calculated as stated in the table. $|G_L|$ and $|G_H|$ is the cardinality of respective clusters. w is defined as shown in the table with M_L and M_H is the mean value of clustered scores. Similarly, percentile pooling also signifies the impact of frames with lower quality scores [13]–[15]. An adjustable parameter κ was used to increase or decrease this impact. P_{low} was defined as the set of frames, which belongs to lower κ_{th} percentile. Primacy pooling take advantage of the tendency of observers, where beginning of the video has a higher impact on overall quality score. Recency effect captures the opposite behaviour, where observers remember the last part of the video while evaluating the overall quality. An adjustable parameter α can be used to increase or decrease these effects [15], [17].

III. Experimental Setup

In the following subsections, we briefly describe the dataset and selected quality metrics, explain the selected temporal subsampling rates, and describe the selected evaluation criteria.

A. Dataset

Extensive experiments were conducted on the vsenseVVDB2 dataset [4] to answer the aforementioned question. This dataset consists of 8 different sequences, see Fig. 2. The sequences were encoded using different compression algorithms at different bitrates, following the MPEG common test conditions. The compression algorithms were G-PCC [18] and V-PCC [18] (both with all-intra and random-access options) for point clouds. These compressed and reconstructed volumetric video sequences were then rendered in Blender using a camera orbiting the 3D model. Then, the subjective quality scores of the rendered sequences, *i.e.*, Mean Opinion Scores (MOS), were collected from a subjective experiment conducted using the absolute category rating with hidden reference (ACR-HR) methodology with a traditional display. The collected MOS values were also made available along with the point cloud sequences. Interested readers are recommended to refer to the original paper [4] for more details.

In our study, we compute both point-based and image-based quality metrics for the evaluation of volumetric videos. Point-based metrics are computed using both the reference and

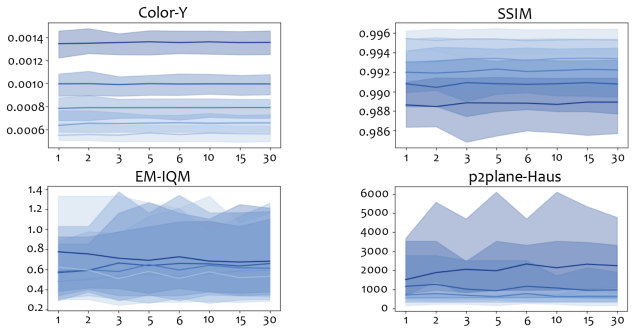


Fig. 3. 95 % percentile range and the median values of the selected metric scores for 5 levels of V-PCC coding at different frame rates.

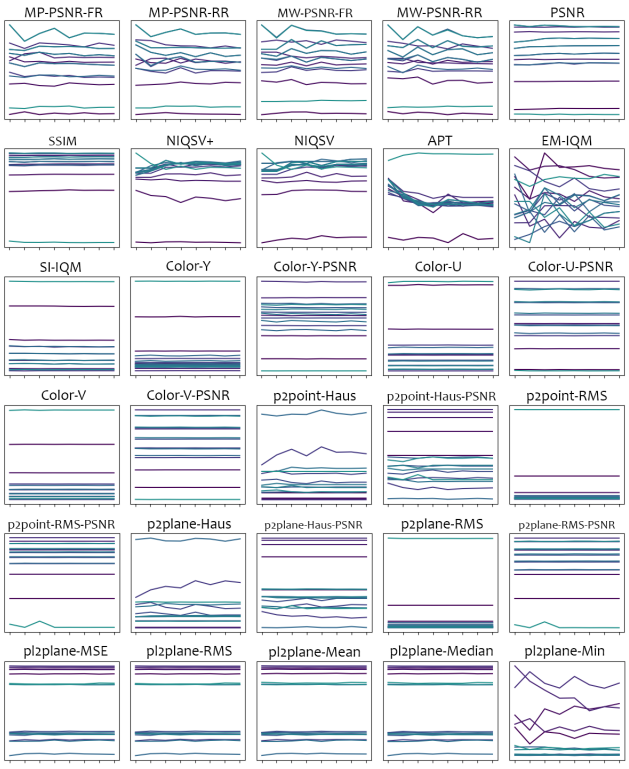


Fig. 4. Each line represents the median metric score that changes over 8 temporal sampling frequencies for a compression type/level over the "AxeGuy" source content. X axis is the fps value for each temporal sampling frequency. Y axis is the metric scores normalized for each metric individually.

the reconstructed volumetric video sequences represented as colored point clouds. Image-based approach uses traditional image quality metrics on the rendered versions of volumetric videos. For this, we employed the same renderings that are used for the subjective quality experiment.

B. Selected quality metrics

We experimented with 11 image-based and 19 point-based quality metrics. Image-based metrics include PSNR, SSIM [19], and 9 other metrics. In [20], Sandic-Stankovic et al proposed MP-PSNR based on multi-scaled pyramid decomposition. Mean Square Error (MSE) was used to quantify the distortion between the reference and distorted image in

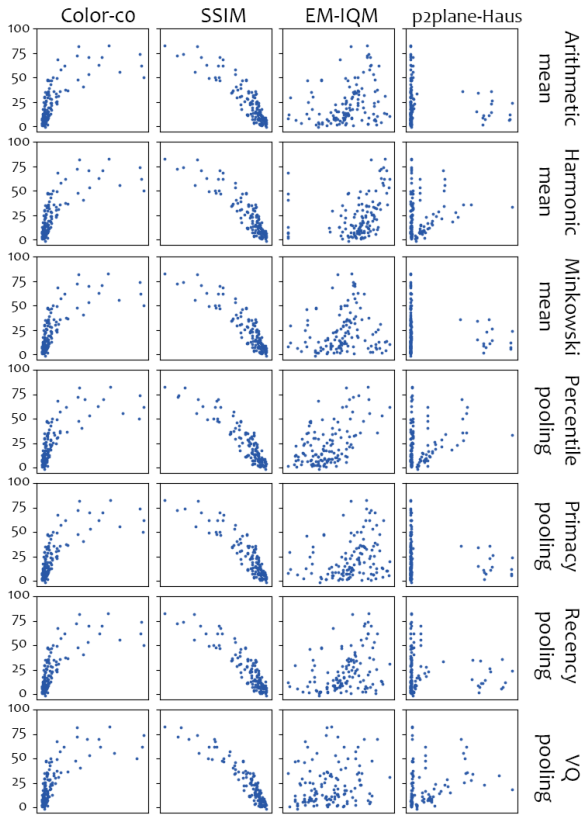


Fig. 5. Scatter plots of objective scores predicted by selected quality metrics versus the DMOS scores. Each row corresponds to a certain pooling method.

sub-bands. Similarly in [21], MW-PSNR was proposed based on morphological wavelet decomposition. Multi-scale wavelet MSE was further utilized to calculate final image quality. Later, reduced versions (RR) of the MP-PSNR and MW-PSNR was proposed in [22]. RR versions use only detail features from higher scales of the decomposition pyramid. By assuming the geometrically relevant distortion is the dominant degradation in free-viewpoint videos, the EM-IQM was presented in [23] to calibrate the structural-deformations. Analogously, to quantify the structural distortions from a higher semantic level, the SI-IQM was proposed [24]. NIQSV [25] was introduced by Tian *et al.* by quantifying non-smooth regions via morphological operations. Later on, it was extended as NIQSV+ [26] by incorporating an extra indicator for dis-occlusion areas. A learning-based blind metric APT [27] was developed by adopting an auto-regression (AR) descriptor.

Point-based metrics considered in this study are based on three main approaches: point-to-point [9], point-to-plane [10], and plane-to-plane [11] differences in 3D space. The term "plane" refers to the plane of a point defined by its normal vector. The missing point normals were estimated using Matlab's `pcnormals` function. The geometry metrics are computed using either root mean square (RMS) distance, mean square error (MSE), or Hausdorff distance measures. Minimum, mean, and median are also used to pool the difference

TABLE II
PCC VALUES BETWEEN METRIC SCORES AND DMOS FOR DIFFERENT TEMPORAL SAMPLING RATES WITH ARITHMETIC MEAN.

	1-fps	2-fps	3-fps	5-fps	6-fps	10-fps	15-fps	30-fps
MP-PSNR-FR	0.7473	0.2655	0.7249	0.2905	0.3680	0.4008	0.3759	0.3088
MP-PSNR-RR	0.7287	0.7063	0.7394	0.7580	0.7098	0.7595	0.7089	0.7594
MW-PSNR-FR	0.7326	0.7432	0.6929	0.7357	0.7413	0.7404	0.7118	0.7378
MW-PSNR-RR	0.7155	0.7596	0.7180	0.7541	0.7269	0.7219	0.7559	0.7591
PSNR	0.8413	0.8298	0.8425	0.8406	0.8305	0.8298	0.8290	0.8286
SSIM	**0.9109**	**0.9093**	**0.9088**	**0.9076**	**0.9083**	**0.9082**	**0.9078**	**0.9081**
NIQSV	0.1806	0.2836	0.1668	0.1548	0.3423	0.3519	0.1596	0.1496
NIQSV+	0.2526	0.2724	0.2857	0.2884	0.2747	0.2727	0.2884	0.2734
APT	0.3624	0.2989	0.3152	0.3125	0.3079	0.3082	0.3088	0.3073
EM-IQM	0.3543	0.3882	0.4035	0.4253	0.4696	0.4309	0.4158	0.4283
SI-IQM	0.8897	0.8872	0.8855	0.8865	0.8861	0.8868	0.8875	0.8871
Color-Y	**0.8498**	**0.8485**	**0.8474**	**0.8464**	**0.8449**	**0.8450**	**0.8472**	**0.8453**
Color-Y-PSNR	0.8495	0.8442	0.8348	0.8460	0.8446	0.8377	0.8464	0.8447
Color-U	0.5552	0.5523	0.5534	0.5545	0.5504	0.5519	0.5548	0.5520
Color-U-PSNR	0.5463	0.5548	0.5572	0.5773	0.5525	0.5457	0.5564	0.5536
Color-V	0.5811	0.5775	0.5819	0.5828	0.5782	0.5781	0.5829	0.5783
Color-V-PSNR	0.5285	0.5771	0.5281	0.5284	0.5559	0.5706	0.5789	0.5688
p2point-Haus	0.2289	0.1649	0.1846	0.3340	0.2059	0.1539	0.1495	0.2092
p2point-Haus-PSNR	0.4670	0.4792	0.4871	0.4796	0.4918	0.4797	0.4875	0.4867
p2point-RMS	**0.9079**	**0.9073**	**0.9081**	**0.9081**	**0.9077**	**0.9068**	**0.9066**	**0.9068**
p2point-RMS-PSNR	0.8743	0.8865	0.8817	0.8795	0.8775	0.8801	0.8789	0.8750
p2plane-Haus	0.1834	0.1150	0.1438	0.1821	0.1224	0.0663	0.1650	0.1274
p2plane-Haus-PSNR	0.2041	0.2041	0.2001	0.1966	0.1898	0.1972	0.1948	0.1994
p2plane-RMS	0.4283	0.4316	0.4314	0.4294	0.4329	0.4433	0.4262	0.4315
p2plane-RMS-PSNR	0.8085	0.8144	0.8123	0.8123	0.5608	0.8130	0.8118	0.8088
pl2plane-MSE	0.4862	0.4813	0.4850	0.4852	0.4862	0.4849	0.4839	0.4863
pl2plane-RMS	0.4870	0.4835	0.4859	0.4873	0.4867	0.4861	0.4870	0.4870
pl2plane-Mean	0.4876	0.4852	0.4879	0.4891	0.4892	0.4884	0.4891	0.4893
pl2plane-Median	0.4876	0.4852	0.4879	0.4891	0.4892	0.4884	0.4891	0.4893
pl2plane-Min	0.0737	0.1434	0.1674	0.1800	0.1690	0.1370	0.1347	0.1282

TABLE III
PCC VALUES BETWEEN METRIC SCORES AND DMOS FOR DIFFERENT POOLING METHODS WITH 30 FPS.

	Arithmetic mean	Harmonic mean	Minkowski mean	Percentile pooling	Primacy pooling	Recency pooling	VQ pooling
MP-PSNR-FR	0.3088	0.7738	0.6641	0.5177	0.3282	0.2996	0.8260
MP-PSNR-RR	0.7594	0.7000	0.7669	0.5283	0.7355	0.7176	0.7932
MW-PSNR-FR	0.7378	0.6817	0.7401	0.5415	0.7010	0.7121	0.7631
MW-PSNR-RR	0.7591	0.7084	0.7251	0.5409	0.7668	0.7291	0.7698
PSNR	0.8286	0.8332	0.8426	0.8103	0.8258	0.8466	0.8428
SSIM	**0.9081**	**0.9081**	**0.9081**	**0.9190**	**0.9077**	**0.9115**	**0.9427**
NIQSV	0.1496	0.1564	0.3436	0.2367	0.3207	0.1561	0.1639
NIQSV+	0.2734	0.2754	0.2745	0.2323	0.2892	0.3154	0.3527
APT	0.3073	0.3073	0.3073	0.3086	0.3052	0.3183	0.1405
EM-IQM	0.4283	0.6565	0.4463	0.6035	0.3913	0.4249	0.1877
SI-IQM	0.8871	0.8874	0.8868	0.8870	0.8892	0.8853	0.8736
Color-Y	**0.8453**	**0.8460**	0.8443	**0.8531**	**0.8445**	**0.8461**	**0.8408**
Color-Y-PSNR	0.8447	0.8448	**0.8449**	0.8296	0.8443	0.8345	0.8352
Color-U	0.5520	0.5556	0.5482	0.5645	0.5520	0.5537	0.5694
Color-U-PSNR	0.5536	0.5533	0.5583	0.5270	0.5432	0.5668	0.5545
Color-V	0.5783	0.5803	0.5766	0.5889	0.5779	0.5777	0.5940
Color-V-PSNR	0.5688	0.5282	0.5800	0.5567	0.5245	0.5842	0.5776
p2point-Haus	0.2092	0.2888	0.1356	0.5604	0.1367	0.1866	-
p2point-Haus-PSNR	0.4867	0.4465	0.3558	0.4806	0.4745	0.4917	-
p2point-RMS	**0.9068**	**0.9050**	**0.9009**	0.9005	**0.9043**	**0.9060**	**0.9021**
p2point-RMS-PSNR	0.8750	0.8848	0.8667	**0.9202**	0.8777	0.8709	0.8184
p2plane-Haus	0.1274	0.1802	0.1234	0.2114	0.1388	0.0992	-
p2plane-Haus-PSNR	0.1994	0.1976	0.0241	0.1455	0.2017	0.1976	0.1028
p2plane-RMS	0.4315	0.4273	0.4300	0.4285	0.4258	0.4447	0.4318
p2plane-RMS-PSNR	0.8088	0.8154	0.7320	0.7648	0.8119	0.8039	0.8886
pl2plane-MSE	0.4863	0.4863	0.4861	0.4613	0.4821	0.4962	0.4670
pl2plane-RMS	0.4870	0.4870	0.4870	0.4751	0.4825	0.4985	0.4678
pl2plane-Mean	0.4893	0.4893	0.4895	0.4774	0.4843	0.4892	0.4539
pl2plane-Median	0.4893	0.4893	0.4895	0.4774	0.4843	0.4892	0.4539
pl2plane-Min	0.1282	0.0842	0.1896	0.1102	0.1792	0.1750	0.2079

scores. In addition to geometry difference, color differences are also calculated using point-to-point correspondence. For this, MSE or PSNR is calculated from the differences between the corresponding points' assigned color values. These color metrics are calculated for Y, U, and V channels.

C. Temporal sub-sampling rates

The frame-rate of the sequences in the vsenseVVDB2 dataset is 30 (fps). For a frame-rate k, we take the first frame among $30/k$ frames and skipped the rest, where the divisor $k \in K = \{1, 2, 3, 5, 6, 10, 15, 30\}$, see Fig. 1.b. This allowed us to uniformly sub-sample the videos.

D. Evaluation criteria

To evaluate the effect of temporal sub-sampling on volumetric video quality assessment accuracy, this paper uses the Pearson correlation coefficient (PCC) and root-mean squared error (RMSE) for prediction accuracy and Spearman rank-ordered correlation coefficient (SROCC) for prediction monotonicity as recommended in [28]. Predicted scores are mapped with a 5-parameter logistic function before evaluation as recommended in [28]. Due to page limits we only report PCC values in the paper. SROCC and RMSE values can be found in the supplementary materials.

IV. RESULTS

The impact of temporal sub-sampling rate: Speeding up the computation of volumetric video quality assessment can be achieved via reducing temporal sampling frequency. To verify whether this can be achieved without sacrificing accuracy, we uniformly sampled volumetric videos from the vsenseVVDB2 dataset with 8 different temporal frequencies. In Fig. 3, 95% percentile ranges for the selected metric scores are presented, where the horizontal axis is the temporal sampling frequency in terms of frame per second. Each line corresponds to the stimuli "AxeGuy" compressed with V-PCC coding [18] at a certain level. While for higher performing metrics such as Color-Y and SSIM, the range of the metric score does not change along with the temporal sampling frequency; metrics such as EM-IQM and p2plane-Haus score fluctuate with the varied temporal sampling frequency. Similarly, Fig. 4 presents all the median metric scores for 16 different compressed versions of "AxeGuy" content. It can be observed that the majority of the metric scores are not affected by temporal sampling frequency. Therefore, metric performances are not significantly affected either. Table II presents the PCC values for each metric under different temporal sampling frequencies. It could be observed that metrics with higher performance (with PCC values higher than 0.5) have insignificant performance differences with varied temporal sampling frequencies.

The impact of temporal pooling method: Fig. 5 shows the scatter plots of the 128 point cloud stimuli in vsenseVVDB2 dataset. The horizontal axis in each plot is the metric score, while the vertical axis is the difference mean opinion score (DMOS) for each stimuli. In the figure, each column corresponds to a different objective quality metric and each row corresponds to a certain temporal pooling method. As shown, for the 1^{st} and 2^{nd} columns (*i.e.,* Color-Y and SSIM), the distributions of the data points of the same column do not differ from each other significantly. On the contrary, the ones of the 3^{rd} and 4^{th} columns vary significantly across different temporal pooling methods. Table III presents the PCC values for each metric with different pooling methods. In the table, it is obvious that the change of temporal pooling method does not significantly affect the high-performing quality metrics (with PCC values higher than 0.5).

Combined effect of temporal sub-sampling and pooling methods: No particular pattern was observed in combined effect evaluation. We further confirmed that no pooling method has clear advantage over certain sub-sampling frequency. We provide a figure reporting the overall accuracy of each pooling

method for each sub-sampling rate in the supplementary materials.

V. Conclusion

In this study, we conducted comprehensive experiments with 30 different metrics to investigate the effect of temporal sub-sampling and temporal pooling methods on the accuracy of volumetric video quality assessment. First, we investigated the effect of the temporal sampling rate. Our findings indicate that, even by sub-sampling the frame rate to 1 fps, metric scores and the metrics' performances do not show a significant difference compared to the full frame rate, i.e., 30 fps. In our experiment with different temporal pooling methods, we observed that better performances were achieved for image-based metrics by using the VQ-Pooling. We did not observe any categorical preference for color and point-based metrics among the tested temporal pooling methods.

Results show the temporal sub-sampling has minimal effect on metrics' correlations with ground truth subjective scores. This observation indicates that compression artifacts affect the perceived quality of the volumetric video uniformly in time. Our findings suggest that with no significant loss in the accuracy of both types of objective quality metrics, calculations can be sped up to 30 times for stimuli with point cloud compression artifacts. It should be noted that further research is required to further extend current conclusions for other types of distortions.

Each considered pooling method has a different priority for the temporal dimension. In our experiments, we observed minimal changes in metric performances with different pooling methods. Similar to the sub-sampling experiments, this indicates the uniform impact of the point cloud compression artifacts on perceived quality.

Our results provide insight regarding performances of various objective metrics for quality evaluation of point cloud compression algorithms on volumetric videos. Additionally, we provide statistical analysis for temporal pooling method selection for each metric. Finally, we show that the objective evaluation of the point cloud compression is minimally affected by the temporal sub-sampling rate, which allows the community to increase the computation efficiency of objective quality evaluation without sacrificing accuracy.

References

[1] A. Smolic, "3D video and free viewpoint video—from capture to display," *Pattern Recognition*, vol. 44, no. 9, pp. 1958 – 1968, 2011. Computer Analysis of Images and Patterns.

[2] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, and S. Sullivan, "High-quality streamable free-viewpoint video," *ACM Trans. Graphics*, vol. 34, July 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. Visual Commun. Image Represent.*, vol. 53, 2018.

[4] E. Zerman, C. Ozcinar, P. Gao, and A. Smolic, "Textured mesh vs coloured point cloud: A subjective study for volumetric video compression," in *12th International Conference on Quality of Multimedia Experience (QoMEX)*, IEEE, May 2020.

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

[6] E. Zerman, P. Gao, C. Ozcinar, and A. Smolic, "Subjective and objective quality assessment for volumetric video compression," in *IS&T Electronic Imaging, Image Quality and System Performance XVI*, 2019.

[7] K. Cao, Y. Xu, and P. Cosman, "Visual quality of compressed mesh and point cloud sequences," *IEEE Access*, vol. 8, pp. 171203–171217, 2020.

[8] M. Pedersen and J. Y. Hardeberg, "Full-reference image quality metrics: Classification and evaluation," *Foundations and Trends® in Computer Graphics and Vision*, vol. 7, no. 1, pp. 1–80, 2012.

[9] 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.

[10] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *IEEE International Conference on Image Processing (ICIP)*, pp. 3460–3464, Sept 2017.

[11] E. Alexiou and T. Ebrahimi, "Point cloud quality assessment metric based on angular similarity," in *International Conference on Multimedia & Expo (ICME)*, 2018.

[12] 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.

[13] S. Rimac-Drlje, M. Vranjes, and D. Zagar, "Influence of temporal pooling method on the objective video quality evaluation," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pp. 1–19, 2009.

[14] M. Seufert, M. Slanina, S. Egger, and M. Kottkamp, ""to pool or not to pool": A comparison of temporal pooling methods for HTTP adaptive video streaming," in *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 52–57, 2013.

[15] Z. Tu, C. J. Chen, L. H. Chen, N. Birkbeck, B. Adsumilli, and A. C. Bovik, "A comparative evaluation of temporal pooling methods for blind video quality assessment," in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 141–145, 2020.

[16] J. Park, K. Seshadrinathan, S. Lee, and A. Bovik, "Video quality pooling adaptive to perceptual distortion severity," *IEEE Transactions on Image Processing*, vol. 22, pp. 610–620, 2013.

[17] B. B. Murdock, "The serial position effect of free recall," *Journal of Experimental Psychology*, vol. 64, no. 5, p. 482, 1962.

[18] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, *et al.*, "Emerging MPEG standards for point cloud compression," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 9, pp. 133–148, Mar 2019.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[20] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "Multi-scale synthesized view assessment based on morphological pyramids," *Journal of Electrical Engineering*, vol. 67, pp. 3–11, 01 2016.

[21] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "DIBR synthesized image quality assessment based on morphological wavelets," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, 2015.

[22] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, "DIBR-synthesized image quality assessment based on morphological multi-scale approach," *EURASIP Journal on Image and Video Processing*, vol. 2017, 07 2016.

[23] S. Ling and P. Le Callet, "Image quality assessment for DIBR synthesized views using elastic metric," in *Proceedings of the 25th ACM international conference on Multimedia*, pp. 1157–1163, 2017.

[24] S. Ling and P. Le Callet, "Image quality assessment for free viewpoint video based on mid-level contours feature," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 79–84, IEEE, 2017.

[25] S. Tian, L. Zhang, L. Morin, and O. Deforges, "NIQSV: A no reference image quality assessment metric for 3D synthesized views," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1248–1252, IEEE, 2017.

[26] S. Tian, L. Zhang, L. Morin, and O. Déforges, "NIQSV+: A no-reference synthesized view quality assessment metric," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1652–1664, 2017.

[27] K. Gu, V. Jakhetiya, J.-F. Qiao, X. Li, W. Lin, and D. Thalmann, "Model-based referenceless quality metric of 3D synthesized images using local image description," *IEEE Transactions on Image Processing*, 2017.

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