Voronoi-based Objective Quality Metrics for Omnidirectional Video

Simone Croci , Cagri Ozcinar , Emin Zerman , Julián Cabrera , and Aljosa Smolic

Abstract—Omnidirectional video (ODV) represents one of the latest and most promising trends in immersive media. The success of ODV depends on the ability to deliver high-quality ODV to the viewers. For this reason, new methods are needed to measure ODV quality that takes into account the interactive look around nature and the spherical representation of ODV. In this paper, we study full-reference objective quality metrics for ODV based on typical encoding distortions in adaptive streaming systems, namely, scaling and compression. The contribution of this paper is three-fold. First, we propose new objective metrics that take into account the unique aspects of ODV. The proposed metrics are based on the subdivision of a given ODV into multiple patches using the spherical Voronoi diagram. Second, we introduce a new dataset of 75 impaired ODVs with different resolutions and compression levels, together with the subjective quality scores gathered during an experiment with 21 participants. Third, we evaluate the proposed Voronoi-based objective metrics using our dataset. The evaluation of the proposed objective metrics and the comparison with existing metrics show that the proposed metrics achieve a better correlation with the subjective scores. The ODV dataset together with the subjective quality scores and the code of the proposed quality metrics are available with this paper.

Index Terms—Omnidirectional video, quality assessment, scaling distortion, compression distortion, Voronoi diagram.

I. INTRODUCTION

Recent years have witnessed significant interest in *omnidirectional video* (ODV), also called 360-degree video, thanks to its ability to create an immersive experience. This emerging media format can be captured by $360°$ video camera systems and rendered through head-mounted displays (HMDs) which let the viewers look around a scene from a fixed viewpoint. For compatibility reasons with conventional video delivery pipelines, ODV is projected onto a 2D plane using various projection techniques, *e.g.*, equirectangular projection (ERP) and cubemap projection (CMP), and projected back onto a sphere surface at rendering time.

Quality measurement is important for the diffusion and success of ODV, which requires considering the unique aspects of ODV. First, this media is spherical in nature, but it is stored and transmitted in planar representations, such as ERP and CMP, which introduce distortions. These distortions must be taken into account in the quality evaluation. Second, HMDs allow the user to freely look around a scene [1], but they show only a fraction of the video, called viewport. Therefore, it is necessary to evaluate the quality locally [2].

Given the importance of quality metrics for ODV, and motivated by a further understanding of the perceptual quality of ODV, we propose new objective quality metrics that consider the spherical nature of ODV and its viewing characteristics. The proposed metrics are characterized by discretizing a given ODV into multiple patches using the spherical Voronoi diagram [3], [4], since they incur in a low projection distortion. Specifically, the proposed objective metrics apply existing metrics designed for traditional 2D content to each patch and average the patch quality scores in order to get a global score.

In order to evaluate the proposed metrics, we created a new dataset of 75 impaired ODVs with different resolutions and compression levels, and we conducted a subjective experiment with 21 participants in order to collect subjective scores. The evaluation shows the effectiveness of the proposed metrics. In particular, we demonstrate the high correlation between the proposed objective metrics and the subjective scores of our dataset. Our ODV dataset along with the subjective quality scores and the source-code of the proposed metrics are available with this paper[1]. We expect that the proposed metrics and the provided dataset will be beneficial for future studies in ODV quality assessment, compression, and streaming.

The remainder of this paper is organized as follows. In Sec. II, we summarize the related work. We then introduce the proposed Voronoi-based objective quality metrics in Sec. III. Afterwards, we describe the technical details of the subjective experiment in Sec. IV. Finally, we present the experimental results in Sec. V and conclude the paper in Sec. VI.

II. RELATED WORK

Most objective quality metrics tailored to ODV consider the spherical viewing representation. The Craster parabolic projection PSNR (CPP-PSNR) metric [5], for instance, uses the Craster parabolic projection to obtain low projection distortion. The Spherical PSNR (S-PSNR) [6] considers uniformly sampled points on the sphere. Furthermore, the weighted

---

spherical PSNR (WS-PSNR) [7] uses weights that consider the projection distortion of the pixels in the planar representation.

Existing subjective quality assessment studies, however, reveal different findings regarding the developed objective metrics for ODV. For instance, the subjective quality assessment results in [8] show that the existing metrics for ODV do not achieve a high correlation with the mean opinion score (MOS). Later, a similar study confirmed these findings by conducting a comprehensive subjective experiment study [9]. Also, the work in [10] concluded that the traditional PSNR metric is the most appropriate one among the existing metrics. On the contrary, some recent works [11], [12] found a superior performance of the objective metrics designed for ODV over the metrics designed for traditional content.

The typical encoding pipeline for video delivery systems introduces two types of visual distortions, namely, compression and scaling. In this context, the video multimethod assessment fusion (VMAF) metric [13] was developed for traditional content, achieving the best performance [14], [15]. However, no related work investigated the quality prediction performance of VMAF and other existing metrics in the presence of scaling and compression distortions in ODV. Furthermore, most of the related work studied low resolution content because of the computational complexity of ODV rendering.

Considering that there is a further need for an understanding of ODV quality measurement, in this work, we performed a subjective quality assessment experiment with our dataset containing high and low resolution ODVs characterized by compression and scaling distortions. Futhermore, we evaluate existing objective quality metrics based on our dataset, and we propose new objective metrics.

III. VORONOI-BASED QUALITY METRICS

This section introduces the proposed objective metrics based on the subdivision of the ODV into Voronoi patches that are analyzed separately.

A. Voronoi Patches

We divide a given ODV into M patches using the spherical Voronoi diagram [3] of M evenly distributed points on the sphere [4], as illustrated in Fig. 1. The computation of M evenly distributed points $\mathbf{P}_i = (X_i, Y_i, Z_i)$ with $i = 0 \ldots M - 1$ on the sphere is obtained according to the following equations: $\alpha_i = i\pi \ (3 - \sqrt{5}$, $Z_i = (1 - 1/n)$ $(1 - 2i/(n-1))$, $d_i = \sqrt{1 - Z_i^2}$, $X_i = d_i \ cos(\alpha_i)$ and $Y_i = d_i \ sin(\alpha_i)$, where α_i is the azimuthal angle and d_i is the distance of the point from the z-axis.

The spherical Voronoi diagram defines the region Π_i for each input point $\mathbf{P}_i$ on the surface of the sphere Ω_S that contains all points which are closer to the $\mathbf{P}_i$ than to any of the other points $\mathbf{P}_l$:

$$\Pi_i = \{\mathbf{P} \in \Omega_S \mid d_S(\mathbf{P}, \mathbf{P}_i) \leq d_S(\mathbf{P}, \mathbf{P}_l) \ \forall l \neq i\}, \quad (1)$$

where $d_S(\mathbf{P}, \mathbf{P}_i)$ is the spherical distance between the current point $\mathbf{P}$ and the point $\mathbf{P}_i$, i.e., the length of the shortest path on the surface of the sphere connecting these two points.
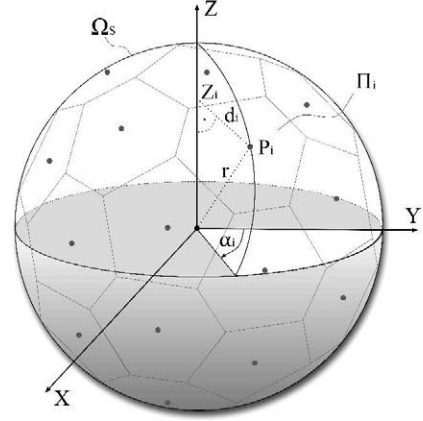

Fig. 1: Spherical Voronoi diagram.

The regions Π_i correspond to the patches that are analyzed separately.

B. Voronoi-based Objective Metrics

Here, we extend existing planar objective metrics to ODV using Voronoi patches. In particular, the following four planar metrics originally developed for traditional content: PSNR, SSIM [16], MS-SSIM [17], and VMAF [13], are extended to VI-PSNR, VI-SSIM, VI-MS-SSIM, and VI-VMAF. For this, we first extract M planar patches using the approach based on the spherical Voronoi diagram. Then, a planar metric for traditional content is applied to each patch, and finally, the patch metric scores are averaged using the arithmetic mean.

To apply the planar metrics, the planar patches Π_i' corresponding to the spherical patches Π_i must be extracted from the ODV. This operation is obtained by first positioning the patch Π_i' plane on the centroid of Π_i, tangent to the sphere. The points on the sphere and the planar patch Π_i' are related by central projection, and the pixel values of the planar patch Π_i' are computed by sampling the ODV in ERP using bilinear interpolation. The resolution of each patch Π_i' is defined by the pixels per visual angle, a parameter that is kept constant for each patch.

IV. SUBJECTIVE EXPERIMENT

Next, we describe the technical details of the conducted subjective experiment, where the quality of different ODVs was evaluated by the participants.

A. Material

We used the following six *uncompressed* and *diverse* ODVs from the joint video exploration team (JVET) of ITU-T VCEG and ISO/IEC MPEG: $\mathcal{V} = \{$*Basketball, Harbor, JamSession, KiteFlite, Dancing, Train*$\}$ [18]–[20]. The *Train* sequence was used as training material to familiarize the participants with the experiment while the other five ODVs were used in the subjective assessment. Each ODV is in 8K 4K ERP and YUV420p format, and of 10 *sec.* length. Spatial and temporal perceptual information measures, SI and TI, were calculated as described in ITU-R P.910 [21], exhibiting diverse

content complexities, as shown in Fig. 2 together with sample thumbnail frames.

B. Design

To investigate the impact of the scaling and compression distortions for ODV, we tested three different resolution sizes at different target bitrate levels. For this purpose, each ODV was downsampled to three different resolution sizes, namely, 2032 × 1016, 3600 × 1800, and 8128 × 4064, before compression using the bi-cubic scaling algorithm of the FFmpeg software (ver. 4.0.3-1 18.04) [22]. Each ODV with a different resolution was then compressed at five different target bitrates, which were selected in a pilot test with experts. In this pre-study, we used the optimal resolution of ODVs displayed by the HMD, 3600 × 1800, as recommended by Zhang et al. [11], and for each tested ODV we decided the target bitrate levels to correspond to different quality levels in the absolute category rating quality scale: ("bad", "poor", "fair", "good", "excellent") [21].

Each downsampled ODV was encoded using the HEVC/H.265 [23] video coding standard. For this, we used the libx265 codec (ver. 2.9) in FFmpeg with the video buffering verifier method to set the target bitrates. We followed the specifications recommended by streaming providers [24], and encode each ODV using two-pass encoding with 150 percent constrained variable bitrate configuration to ensure smooth perceptual video quality frame by frame. We also defined the buffer size during encoding which limits the output bitrate to two times the maximum bitrate to handle large bitrate spikes.

Before the subjective experiment, each stimulus was up-sampled to the resolution of 8128 × 4064 using the bi-cubic scaling algorithm of FFmpeg after decoding. The reason for this up-sampling was to eliminate the impact of the unknown resampling algorithm of the used video player by playing the highest possible resolution that the player can render ODV for the HMD.

C. Apparatus

In the subjective experiment, each stimulus was shown to the participants using the HTC Vive HMD and played with the Virtual Desktop application. Virtual Desktop is an ODV player and an application that enables the users to watch and interact with the desktop using the HMD and VR controllers. We also used an open-source MATLAB GUI presented in [25], [26]. This GUI allows the participants to assign a quality opinion score to each stimulus without taking off the HMD. Additionally, we recorded the participants' HMD viewing directions while they were watching the stimuli.

D. Procedure

To perform the subjective study, we followed the modified-absolute category rating (M-ACR) [27] procedure. We chose the M-ACR, because it was demonstrated in recent evaluations [27], [28] that it is more reliable than existing methods developed for traditional 2D video. This methodology increases the duration of exposition time by showing each stimulus twice with three seconds of a mid-gray screen in between these two presentations. Afterwards, the participants were able to assign their opinion quality score to the stimulus.

Participants were seated in a swivel chair and allowed to turn freely. There were two sessions; the first session was split into a training and a test phase, while the second session had only a test phase. Each session lasted less than half an hour. During the training phase, the Train video sequence with five different quality levels was displayed. Then, during the test phase, impaired ODVs were randomly displayed (avoiding consecutive representation of the same content) while the individual viewport trajectories were recorded. Furthermore, the quality scores were assigned by the participants based on a continuous grading scale in the range [0,100], where the best grade corresponds to 100, as recommended in ITU-R BT.500-13 [29].

E. Participants

In all, 24 participants, 20 males and four females, took part in our subjective experiment. Participants were aged between 22 and 38 with an average of 29.71 years. The gathered quality opinion scores were screened for outliers using the outlier detection method recommended in ITU-R BT.500-13 [29]. Three outliers were found and removed. All participants were screened for visual acuity and found to have normal or corrected-to-normal vision.

F. Data Processing

From the participants' raw subjective scores, we compute the difference mean opinion score (DMOS). For this task, we apply the standard approach described in [30]. Given s_{ij} and s_{ij}^r as the raw subjective scores assigned by participant i to the impaired ODV j and the corresponding reference ODV, respectively, we first compute the difference score as: $d_{ij} = s_{ij}^r - s_{ij}$. Next, the difference score d_{ij} is converted to the z-score as follows: $z_{ij} = (d_{ij} - \mu_i)/\sigma_i$, where μ_i and σ_i are the mean and standard deviation of the scores assigned by the participant i. Afterwards, as recommended in ITU-R BT.500-13 [29], a participant is rejected, if 5% of her/his z-scores are outside the range of two standard deviations from the mean z-scores. In the next step, the z-scores are linearly rescaled in the interval [0,100] as follows: $z_{ij}' = 100(z_{ij}+3)/6$. The rescaling is based on the assumption that the z-scores z_{ij} are normally distributed with mean equal to zero and standard deviation equal to one, which means, that 99% of the z-scores z_{ij} are in the interval [-3,3], and consequently 99% of the rescaled z-scores z_{ij}' are in the interval [0,100]. The final $DMOS_j$ value of ODV j is then obtained by averaging the rescaled z-scores z_{ij}' of the N participants excluding the outliers:
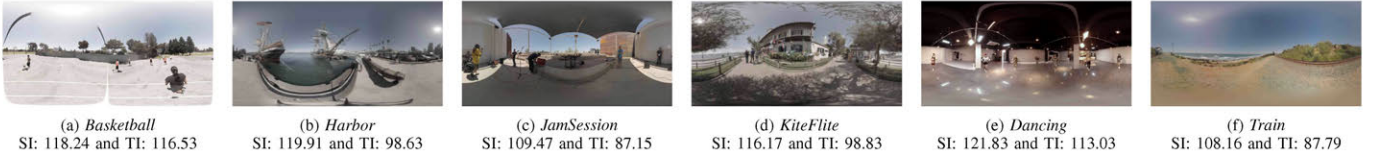
$$DMOS_j = \frac{1}{N} \sum_{i=1}^{N} z_{ij}'. \tag{2}$$

Fig. 2: Sample thumbnail frames of the six ODVs with their SI and TI measures used in the subjective experiment.

(a) *Basketball*   (b) *Harbor*   (c) *JamSession*

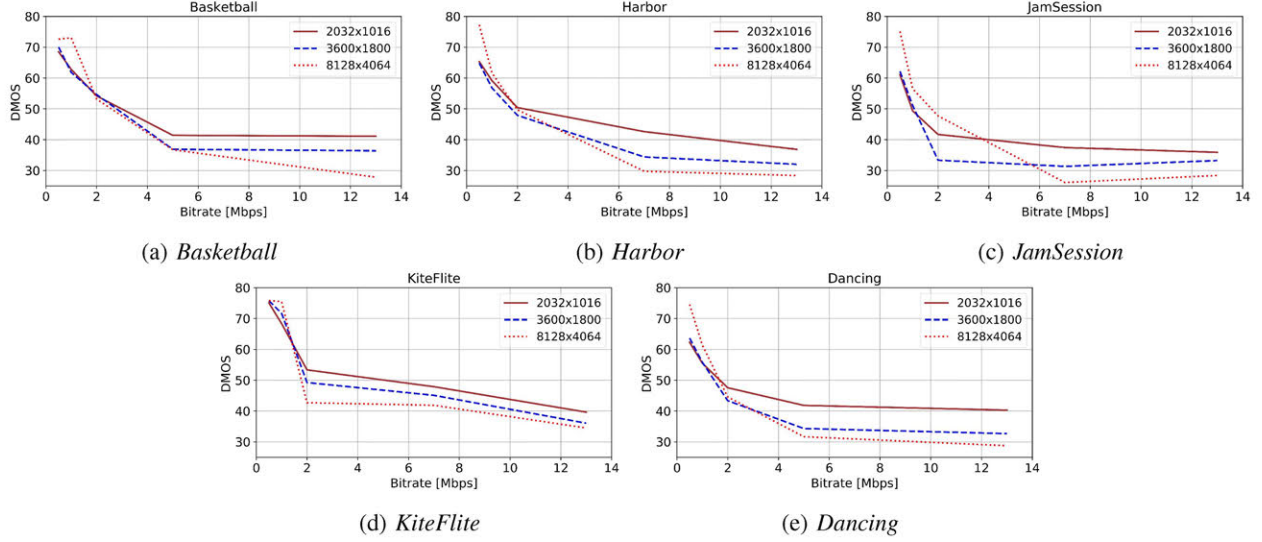
(d) *KiteFlite*   (e) *Dancing*

Fig. 3: DMOS values of the ODVs used in the subjective experiment.

V. RESULTS

In the following section, we investigate the subjective data gathered during our experiment, and we evaluate the performance of a selection of existing objective metrics and our proposed Voronoi-based quality metrics.

A. Evaluation of the Subjective Data

Fig. 3 shows the plots of the DMOS values of the experiment ODVs computed based on the approach described in Section IV-F. Looking at the figure, we noticed that the highest resolution videos (8128 × 4064) have the best quality (lowest DMOS value) at the highest bitrate (13 *Mbps*), while the lowest resolution videos (2032 × 1016) have the best quality at the lowest bitrate (500 Kbps). These findings are especially important for ODV adaptive streaming systems [31], where the selection of the optimal encoding parameters is crucial.

B. Evaluation of Objective Metrics

1) Existing metrics: To evaluate the performance of the objective metrics, we analyzed the correlation between the subjective and objective scores by fitting a logistic function to map the objective scores to the subjective scores. For this, we used the logistic function in [32] as defined by

$$s' = \frac{\beta_1 - \beta_2}{1 + e^{-\frac{S - \beta_3}{\|\beta_4\|}}} + \beta_2, \qquad (3)$$

where s' is the predicted subjective score of the objective score s. In our evaluation of the objective metrics, the subjective score predicted by the logistic function is the reversed DMOS (*i.e.*, subtracted from 100).

To compare the objective metrics, the following performance metrics were applied in order to evaluate how well the logistic function predicts the subjective score: Pearson's linear correlation coefficient (PLCC), Spearman's rank ordered correlation coefficient (SROCC), root mean squared prediction error (RMSE), and mean absolute prediction error (MAE). PLCC and SROCC measure the prediction accuracy and the monotonicity, respectively. The larger these two metrics are, the more accurate and monotonic the prediction is. For RMSE, and MAE, the smaller the metric, the better the performance of the prediction is.

We evaluated the performance of eight objective quality metrics applied to the luminance channel of the ODVs used in our subjective experiment. The first four selected metrics were originally developed for traditional content, namely, PSNR, SSIM [16], MS-SSIM [17], and VMAF [13]. The video quality measurement tool [33] and the official code provided by Netflix [34] were used for the first three metrics and VMAF, respectively. The first three metrics are widely used for image and video quality evaluations, while VMAF is the best performing metric for traditional video. These metrics were applied to ODVs in two different formats, ERP and CMP. The remaining four metrics that we evaluated were explicitly designed for ODV, namely, S-PSNR-I (S-PSNR with pixel interpolation), S-PSNR-NN (S-PSNR without pixel interpolation), WS-PSNR, and CPP-PSNR. For these metrics, we used the implementation of 360Lib [35].

Table I shows the performance of the eight selected objective metrics. Looking at the Table, we can notice a small

TABLE I: Performance evaluation of the selected existing objective metrics and the new Voronoi-based metrics together with two projection formats. The best performance values are in bold.

Metrics	Representation	PLCC	SROCC	RMSE	MAE
PSNR	ERP	0.8292	0.7979	8.7921	7.0102
PSNR	CMP	0.8429	0.8101	8.4822	6.7224
S-PSNR-I	ERP	0.8479	0.8139	8.3675	6.5937
S-PSNR-NN	ERP	0.8489	0.8150	8.3432	6.5718
WS-PSNR	ERP	0.8485	0.8141	8.3519	6.5790
CPP-PSNR	ERP	0.8479	0.8136	8.3690	6.5954
SSIM	ERP	0.7347	0.7107	10.5253	8.5131
SSIM	CMP	0.7419	0.7209	10.4370	8.5427
MS-SSIM	ERP	0.9085	0.8888	6.6162	5.3242
MS-SSIM	CMP	0.9125	0.8954	6.4904	5.1064
VMAF	ERP	0.9160	0.8861	6.2562	4.7724
VMAF	CMP	0.9267	0.8998	5.9792	4.4919
VI-PSNR	ERP	0.8545	0.8251	8.1746	6.4750
VI-SSIM	ERP	0.8132	0.7968	9.1138	7.2579
VI-MS-SSIM	ERP	0.9447	0.9334	5.2625	4.2398
VI-VMAF	ERP	**0.9661**	**0.9499**	**4.2356**	**3.1269**

improvement when the metrics are applied to the CMP format instead of the ERP format. Moreover, the performance of the PSNR based metrics developed for ODV is close to the performance of the standard PSNR applied to the CMP format. Furthermore, among the evaluated metrics SSIM is characterized by the worst performance, even worse than PSNR. On the other hand, the metrics with the best performance are MS-SSIM and VMAF. Between these two metrics, VMAF is slightly better than MS-SSIM for both representation formats.

2) Proposed Voronoi-based metrics: This section presents the evaluation of the proposed metrics based on the Voronoi patches. For the extraction of the Voronoi planar patches from the ODV, we selected an angular resolution equal to 15 pixels per degree, a resolution close to the one of the HMD used in our subjective experiment (HTC Vive). The number of Voronoi patches used in this evaluation is 15. We evaluated the metrics also with 10 and 20 Voronoi patches without noticing large differences in the performance. The performance of these metrics is presented in Table I. According to the values in the table, the new Voronoi-based metrics are characterized by a better performance than the original planar metrics applied to the ERP and CMP formats. This is expected because of the lower projection distortion of the Voronoi patches compared to the ERP and the CMP. Among the new metrics reported in the table, the worst performing one is VI-SSIM. This is not unexpected, since SSIM is also the worst performing metric when applied to the ERP and CMP formats. After VI-SSIM, the next better performing metric is VI-PSNR, which has a comparable performance to the PSNR-based metrics developed for ODV. The best performing metric is VI-VMAF followed by VI-MS-SSIM that achieves a slightly worse performance.

Fig. 4 shows a comparison between the fitted logistic functions of PSNR in ERP format and VI-VMAF. In these plots, the higher correlation of VI-VMAF with the subjective scores is visible.

Fig. 5 shows the visualization of the VMAF patch scores, where the patch scores are color-coded based on the jet

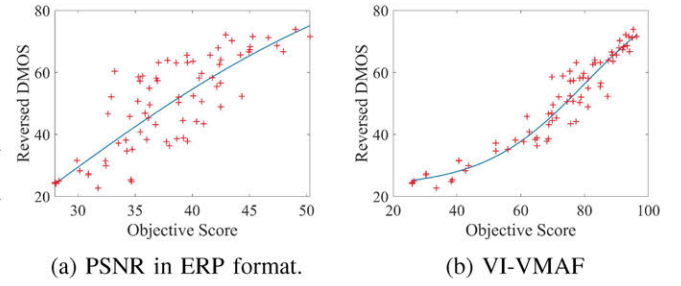(a) PSNR in ERP format.   (b) VI-VMAF

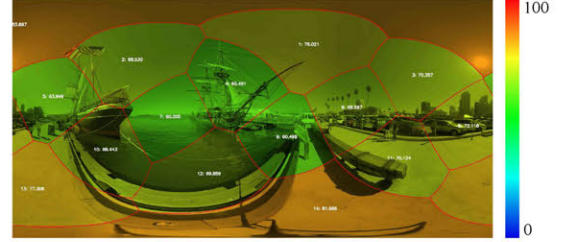Fig. 4: Objective vs subjective score plots with the fitted logistic function.

Fig. 5: Visualization of the VMAF patch scores of the *Harbor* ODV with resolution 2032 × 1016 and encoded with target bitrate equal to 2 Mbps.

colormap and visualized in the Voronoi diagram mapped to the ERP format. This visualization is useful in order to easily identify the quality of the different regions of the ODV. As can be seen in the figure, the patch scores close to the pole caps are larger than the ones close to the equator. This can be explained by the larger pixel density at the pole caps of the ERP format. This visualization also shows that diverse VMAF scores can exist in different regions of the ERP. Hence, the extraction and analysis of multiple planar patches generated by the Voronoi diagram is essential to evaluate the quality locally, substantially improving the performance of the metrics, especially of VMAF.

VI. CONCLUSIONS

In this paper, we studied the objective quality evaluation of ODV. In particular, we proposed new objective metrics based on the subdivision of the ODV into Voronoi patches. These new metrics were evaluated based on subjective data of 75 ODVs with different resolutions and HEVC/H.265 compression levels. In order to gather the subjective data, we conducted a subjective experiment with 21 participants. The new objective metrics were compared with existing metrics applied to two different planar projection formats. This evaluation revealed that the new objective metrics achieve a better performance. Among all the metrics considered in this paper, the one with the best performance is VI-VMAF. The ODV dataset used in our experiment and the related subjective quality scores, together with the code of the new metrics, are made public with this paper.

In the future, we plan to evaluate the new metrics based on videos with different distortions than the resampling and HEVC/H.265 distortions considered in this work. Besides, we

also intend to extend these new metrics by integrating visual attention information.

REFERENCES

[1] C. Ozcinar and A. Smolic, "Visual attention in omnidirectional video for virtual reality applications," in *10th International Conference on Quality of Multimedia Experience (QoMEX 2018)*, Sardinia, Italy, May 2018.

[2] C. Ozcinar, J. Cabrera, and A. Smolic, "Visual attention-aware omnidirectional video streaming using optimal tiles for virtual reality," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2019.

[3] F. Aurenhammer, "Voronoi diagrams - a survey of a fundamental data structure," *ACM Computing Surveys*, vol. 23, no. 3, pp. 345–405, 1991. [Online]. Available: http://portal.acm.org/citation.cfm?doid=116873.116880

[4] S. Croci, S. Knorr, L. Goldmann, and A. Smolic, "A framework for quality control in cinematic vr based on voronoi patches and saliency," in *International Conference on 3D Immersion*, Brussels, Belgium, 2017.

[5] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," *Proc.SPIE*, vol. 9970, pp. 9970 – 9970 – 9, 2016. [Online]. Available: https://doi.org/10.1117/12.2235885

[6] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *2015 IEEE International Symposium on Mixed and Augmented Reality*, Sept 2015, pp. 31–36.

[7] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1408–1412, Sept 2017.

[8] E. Upenik, M. Reábek, and T. Ebrahimi, "A testbed for subjective evaluation of omnidirectional visual content," in *Proceedings of the Picture Coding Symposium (PCS)*, 2016.

[9] E. Upenik, M. Rerabek, and T. Ebrahimi, "On the performance of objective metrics for omnidirectional visual content," *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017.

[10] H. T. T. Tran, N. P. Ngoc, C. M. Bui, M. H. Pham, and T. C. Thang, "An evaluation of quality metrics for 360 videos," in *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, July 2017, pp. 7–11.

[11] Y. Zhang, Y. Wang, F. Liu, Z. Liu, Y. Li, D. Yang, and Z. Chen, "Subjective panoramic video quality assessment database for coding applications," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 461–473, June 2018.

[12] W. Sun, K. Gu, S. Ma, W. Zhu, N. Liu, and G. Zhai, "A Large-Scale compressed 360-degree spherical image database: From subjective quality evaluation to objective model comparison," in *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, Aug. 2018, pp. 1–6.

[13] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652, Jan 2019.

[14] N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, and S. Möller, "An evaluation of video quality assessment metrics for passive gaming video streaming," in *Proceedings of the 23rd Packet Video Workshop*. ACM, 12 Jun. 2018, pp. 7–12.

[15] R. Rassool, "VMAF reproducibility: Validating a perceptual practical video quality metric," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun. 2017, pp. 1–2.

[19] E. Asbun, H. He, H. Y., and Y. Ye, "AHG8: Interdigital test sequences for virtual reality video coding," JTC1/SC29/WG11, ISO/IEC, Chengdu, China, Tech. Rep. JVET-D0039, Oct 2016.

[16] 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, April 2004.

[17] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, Nov 2003, pp. 1398–1402 Vol.2.

[18] A. Abbas and B. Adsumilli, "AHG8: New GoPro test sequences for virtual reality video coding," JTC1/SC29/WG11, ISO/IEC, Chengdu, China, Tech. Rep. JVET-D0026, Oct 2016.

[20] G. Bang, G. Lafruit, and M. Tanimoto, "Description of 360 3D video application exploration experiments on divergent multiview video," JTC1/SC29/WG11, ISO/IEC, Chengdu, China, Tech. Rep. MPEG2015/M16129, Feb. 2016.

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

[22] "x265 HEVC Encoder / H.265 Video Codec," http://x265.org/, Jan 2018.

[23] J.-R. Ohm and G. Sullivan, "Vision, applications and requirements for high efficiency video coding (HEVC)," ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Tech. Rep. MPEG2011/N11891, March 2011.

[24] "HLS Authoring Specification for Apple Devices," https://developer.apple.com, Jan 2018.

[25] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan, "Assessing visual quality of omnidirectional videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, pp. 1–1, 12 2018.

[26] https://github.com/Archer-Tatsu/Evaluation_VR-onebar-vive, accessed: 2019-01-15.

[27] A. Singla, S. Fremerey, W. Robitza, P. Lebreton, and A. Raake, "Comparison of subjective quality evaluation for HEVC encoded omnidirectional videos at different bit-rates for UHD and FHD resolution," in *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, ser. Thematic Workshops '17. New York, NY, USA: ACM, 2017, pp. 511–519. [Online]. Available: http://doi.acm.org/10.1145/3126686.3126768

[28] A. Singla, W. Robitza, and A. Raake, "Comparison of subjective quality evaluation methods for omnidirectional videos with DSIS and modified ACR," in *IS&T Electronic Imaging, Human Vision and Electronic Imaging XXII*. International Society for Optics and Photonics, 2018.

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

[30] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, June 2010.

[31] C. Ozcinar, A. De Abreu, S. Knorr, and A. Smolic, "Estimation of optimal encoding ladders for tiled 360 VR video in adaptive streaming systems," in *The 19th IEEE International Symposium on Multimedia (ISM 2017)*, Taichung, Taiwan, Nov. 2017.

[32] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," ITU, COM 9-80-E, Geneva, Switzerland, Tech. Rep., 2000.

[33] "Video quality measurement tool (VQMT)," https://mmspg.epfl.ch/vqmt, accessed: 2019-01-15.

[34] "Video multi-method assessment fusion (VMAF)," https://github.com/Netflix/vmaf, accessed: 2019-01-15.

[35] "360lib," https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/trunk, accessed: 2019-01-15.