https://doi.org/10.2352/ISSN.2470-1173.2020.9.IQSP-284

Subjective and Viewport-based Objective Quality Assessment of 360-degree videos

Roberto G. de A. Azevedo[1], Neil Birkbeck[2], Ivan Janatra[2], Balu Adsumilli[2], Pascal Frossard[1]
[1]Ecole Polytechnique Fédérale de Lausanne (EPFL); Lausanne, Switzerland.
[2]Youtube, Mountain View, California, USA.

Abstract

Visual distortions in processed 360-degree visual content and consumed through head-mounted displays (HMDs) are perceived very differently when compared to traditional 2D content. To better understand how compression-related artifacts affect the overall perceived quality of 360-degree videos, this paper presents a subjective quality assessment study and analyzes the performance of objective metrics to correlate with the gathered subjective scores. In contrast to previous related work, the proposed study focuses on the equiangular cubemap projection and includes specific visual distortions (blur, blockiness, H.264 compression, and cubemap seams) on both monoscopic and stereoscopic sequences. The objective metrics performance analysis is based on metrics computed in both the projection domain and the viewports, which is closer to what the user sees. The results show that overall objective metrics computed on viewports are more correlated with the subjective scores in our dataset than the same metrics computed in the projection domain. Moreover, the proposed dataset and objective metrics analysis serve as a benchmark for the development of new perception-optimized quality assessment algorithms for 360-degree videos, which is still a largely open research problem.

Introduction

Virtual reality (VR) and immersive media technologies have recently drawn significant interest from both end users and researchers, resulting in many incremental advancements in the last few years. As one of these new immersive media technologies, omnidirectional (or 360-degree) images and videos that capture the 360-degree field of view of a scene, have gained popularity in many areas, such as entertainment, education, and robotics. Omnidirectional visual content provides the user with the ability to navigate the scene with three degrees of freedom (3DoF), resulting in more realistic and immersive experiences, particularly when the user consumes the content wearing a head-mounted display (HMD). Such content, however, requires a massive amount of data to be captured, processed, and transmitted; this challenges current technologies and infrastructure and calls for new solutions in the whole immersive media delivery pipeline.

Despite being naturally a spherical signal, 360-degree signals are commonly mapped to and stored as planar 2D images in order to reuse existing image and video processing technologies [5]. Different projections have been proposed to optimize the pixels distribution on the planar layout, such as equirectangular (ERP), cube map (CMP), and equiangular cube map (EAC)[1] [5]. To be presented to the end user, the planar signal is mapped back to the spherical domain, and, at each instant, the portion of the sphere watched by the viewer, known as *viewport*, is rendered and displayed. When the visualization happens via an HMD, the viewports (one per eye) are seamlessly updated following the user's head movements, which provides the user with an increased sense of presence.

Similar to traditional 2D image and video, visual quality assessment (VQA) methods also play a fundamental role in the different stages of the immersive media distribution chain, e.g., in the design of recording devices, representations, and compression of digital information, as well as streaming, and display. Indeed, advances in immersive media VQA are of fundamental importance to provide high-quality and optimized services that guarantee the best trade-off between resource usage and good end-user quality of experience (QoE). Although it is related to traditional 2D and stereoscopic 3D visual content, VQA of processed 360-degree media visualized via HMDs brings its own specificities. In particular, the appearance of new types of distortions [1], the magnification of the content, an increased field-of-view, and the fact that the user is fully immersed in the content completely change the QoE perspective. These new requirements call for the development of new methods and good practices for the visual quality and QoE assessment of 360-degree visual content [1].

This paper describes a subjective quality study and establishes a dataset to get a deeper understanding on the perceptibility of compression artifacts and how they affect the overall QoE of 360-degree visual content consumed through HMDs. Previously, we have performed a preliminary experiment to get the range of allowed visual distortions in 360-degree videos [2]. Based on those initial findings, we perform here a subjective VQA study based on the Single Stimulus with hidden reference [10] methodology. In total, 21 subjects rated 128 sequences (120 distorted + sequences + 8 original ones), where the included distortions are blur, blockiness, cube map seams [1], and H.264 compression. Finally, we provide the subjective data analysis and measure the performance of standard objective quality metrics computed in both the projection and viewport domains on our dataset. The developed dataset is unique when compared to previous work because: (1) it includes individual visual distortions (previous work on 360-degree images/video subjective VQA include distortions through different quality parameters in traditional compression algorithms, which results in mixed visual distortions); (2) it is the first to use the EAC projection to handle spherical signal; and (3) it mixes both monoscopic and stereoscopic versions of the content.

The rest of the paper is structured as follows. First, we discuss related work and compare them with our study. Then, we detail the methodology, stimulus preparation, and subjects' profile of our subjective study. Based on the data gathered on the

[1]https://youtube-eng.googleblog.com/2017/03/improving-vr-videos.html

subjective study, we present the analysis of objective metrics performance. Finally, conclusions and future work are discussed.

Related work

Even though there is still no standard methodology for performing subjective 360-degree quality studies, researchers agree that it does not come as a straightforward extension of traditional 2D-only studies. The research community has been actively discussing best practices for 360-degree VQA, and some subjective quality studies have been recently performed [1, 13]. Initial experiments were based on presenting the rendered viewports using traditional displays [4, 21]. Such an approach, however, lacks important immersive features that can only be assessed when the user is wearing an HMD. Hence, we focus mainly on the studies in which the user uses an HMD. Most of the methodologies in use today for 360-degree subjective VQA are adaptations of traditional ones, used for 2D videos in planar displays. They include both Single Stimulus —e.g., Absolute Category Rating (ACR) and Absolute Category Rating with Hidden Reference (ACR-HR) [9]— and Double Stimulus methods [10]. When performing subjective tests using HMDs, additional data such as head and eye motion, sickness, and sense of presence are also interesting [1]. The recent studies include both image-only and video-based VQA.

Regarding image datasets, Upenik et al. [17] presents a testbed for the subjective evaluation of 360-degree images that allows both to collect raw scores and record tracking information. Based on the proposed testbed, they discuss a pilot study using 6 different omnidirectional image contents compressed with four JPEG quality parameters. ACR-HR with a five-grade quality scale is used as the test methodology. Based on the testbed proposed in [17], Perrin et al. [14] uses a dual stimulus method for measuring the quality of High Dynamic Range (HDR) omnidirectional images considering well-known tone mapping operators. Huang et al. [8] presents a single stimulus ACR-based study for omnidirectional images. It studies both spatial resolution and JPEG distortions for ERP images. The study found that the ideal viewing duration for 360-degree images is 20 seconds, which allows the user to explore the content entirely.

Regarding video-based datasets, Xu et al. [18] establishes a video quality dataset using the SSCQS (Single Stimulus Continuous Quality Scale) methodology [10]. Before building the dataset, they perform a preliminary study to analyze the users' head motion (HM) when watching 360-degree videos. They conclude that: (i) the viewing directions are highly consistent with the video content, hence, there is no need for fixing the viewing directions in VQA studies for 360-degree content; and (ii) at least 20 subjects should be involved in 360-degree VQA studies. In particular, allowing free head movement during controlled subjective tests has the advantage of keeping the test environment closer to the real one where the media will be consumed. Duan et al. [7] presents a dataset composed of 10 original 360-degree videos, 15 seconds each, which are rated by 13 subjects using a single stimulus 0–5 continuous scale. The study explores the influence of resolution, frame rate, and different H.264 bitrate. Li et al. [12] establishes a database that includes both VQA scores and the corresponding HM and eye motion (EM) data for studying the relationship between visual quality and human behavior. The dataset includes three projections types, ERP, RCMP (Reshaped Cubemap), and TSP (Truncated Square Pyramid), and three H.265 compression levels as distortions. Queluz et al. [15] presents the result of a subjective quality study for ERP 360-degree videos that includes spatial and temporal subsampling, HEVC compression, and their combination. Based on the gathered data, the authors also analyze the performance of different objective metrics on their dataset.

Building on the lessons learned in the previous studies, we report here our subjective VQA study focused on the impact of specific visual distortions on the perceived quality of 360-degree content. Different from previous work, we add specific visual distortions (blur, blockiness, and seams) and mix both monoscopic and stereoscopic versions of the original content in our dataset. By mixing the monoscopic and stereoscopic versions of the content, we can provide insights on how the same artifact affects viewers' perceived quality on mono and stereo viewing modes. Also, to allow us to analyze the impact of mixing difference visual distortions, we include H.264/AVC compression as one of the possible distortions. Finally, we focus on the EAC map projection, which has not been studied by any of the previous works, and analyzes the performance of objective metrics computed in the viewports domain, which is also a novelty of our work.

VQA Subjective study

Environment setup

The subjective experiment was performed in two sites: Youtube, Mountain View, CA; and EPFL, Lausanne, Switzerland. In both sites, we used a Lenovo Mirage Solo HMD (2560x1440@75Hz) for visualization and the user sits on a swivel chair so that she/he is free to explore the content. We developed our own software that allows the user to watch the 360-degree video content and rate it without the need of removing the HMD. Fig. 2 shows the user inferface used to select the quality score.

Methodology

The used methodology is based on Single stimulus with hidden reference [9]. The user evaluates each stimulus, one-by-one, and voting is performed after each viewing. Both the original (pristine) and the distorted content are included so that it is possible to take into account the video content when computing the subjective ratings. Subjects do not know which ones are the reference content. We provide the user with a continuous 0–100 scale divided into a 5-level categorical scale ("5-Excellent", "4-Good", "3-Fair", "2-Poor", and "1-Bad") (see Fig. 2). Each stimulus duration is 10s and the same stimulus is presented twice, with a 2s mid-gray screen between them, so that the user is able to explore different parts of the 360-degree content. Fig. 1 shows the methodology used in the experiments, which is divided into:

Study presentation. The subjects are presented with the goal of the study and informed that, due to characteristics of distortions in 360-degree videos and the use of an HMD, they might feel some discomfort or sickness, in which case, they are free to leave the experiment at any time.

Pre-experiment subject screening. The subjects are screened for color blindness and stereoscopic vision issues. The color blindness test is based on the Ishihara Color Test[2]. For the stereoscopic vision test, we show a stereoscopic 3D video pattern on the HMD. The participants can see different numbers if they have normal stereoscopic vision.

[2]https://colour-blindness.com/colour-blindness-tests/ishihara-colour-test-plates/

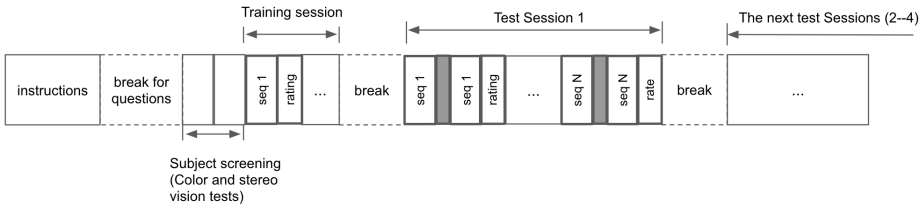


Figure 1: Methodology

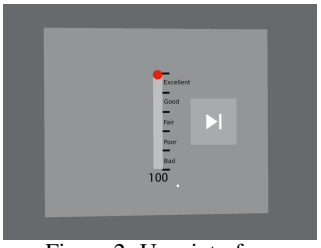


Figure 2: User interface

Training. The user is presented with the overall process. Examples of distorted content containing the minimum, medium, and maximum levels of each distortion are presented to the subjects. Also, they are informed about each of the distortion levels associated with the presented content. The data gathered in this session is not used in the data analysis process.

Test sessions (1–4). The stimuli are randomly divided into 4 equal-sized sets that were evaluated in 4 sessions by each user. In each session, the subject watches each stimulus (pristine or distorted sequence) twice, with the 2s mid-gray screen between them. After that, the subject is asked to rate the quality of that stimulus using the scale depicted in Fig. 2. To avoid fatigue, we perform a break of 5–10min. between the sessions. The subject is able to control the time of the break so that he could continue when she/he is comfortable to do it. Due to time restrictions, some subjects also performed the first two sessions in one day and the next ones in the following day. Besides gathering the subjective score given to each stimulus, we also record the head motion of the user while watching the 360-degree content in our dataset.

Stimuli

Source sequences. 5 original sequences are included in the study (see Fig. 3), being 3 stereoscopic content and 2 monoscopic ones. Based on the original ERP content, we re-project each of the sequences to EAC, using a 3840x2160 resolution. For the original stereoscopic sequences, we added both the monoscopic and the stereoscopic versions to the study.

Distortion simulation. We consider four main distortions: blocking, seams, blur, and H.264 compression. The *blocking* distortions are simulated using the H.263 algorithm with different quality parameters (QP). The *seam* distortion is simulated through a Gaussian blur filter only in the discontinuities (i.e., where the cube faces meet) in the EAC content. The *blur* distortion is added through a Gaussian filter on the whole frame. To avoid the appearance of seams in the blur content, the blur command is performed on a padded version of the EAC frame. Then, the correct resolution is cropped from it. Finally, the *H.264* compression distortions are created using the H.264 algorithm with different QPs. More details on the commands used for generating the different distortions can be found in [2].

Distortion levels. The visibility and range of visual distortion perceptibility for 360-degree content watched on HMDs are not the same as those of traditional images watched on planar 2D displays. Thus, it is not straightforward to use the same distortion levels as the ones commonly used in traditional 2D visual quality studies. Prior to performing our VQA experiments, we have performed a task-based exploratory study to better understand the range of allowed distortions in 360-degree visual content [2]. We have asked the subjects to find, using an interactive interface, the *first just-noticeable distortion* (JND) and the *Break-in-Presence* points when incrementally adding distortions to 360-degree visual content. JND is a statistical quantity that accounts for the maximum distortion that stays unnoticeable to a human being [20]. We define the "1st JND" as the maximum amount of distortion that can be added to the pristine (non-distorted) content before the user perceives any difference in it. Presence is defined as the sense of "being there", inside a space, even when physically located in a different location [11]. We define the Break-in-presence point as the amount of visual distortion that can not anymore induce a sense of presence. The chosen distortion levels are then sampled inside the "1st JND" and Break-in-presence ranges through expert viewing.

Subjects

In total, we recruited 21 subjects to participate in the study, 14 males and 7 females. All the participants have passed on the pre-experiment subject screening tests.

Data processing

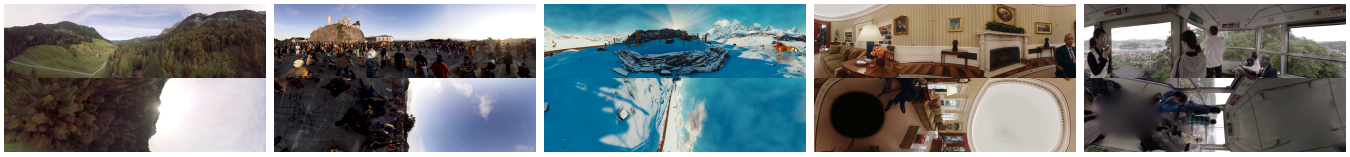
A common issue with subjective studies is that different users use the rating scale differently. To account for such issues, we process the RAW score based on BT.500-12 [10]. First, for stimuli j and subject i, we compute the difference score: $d_{ij} = s_{ij}^{\text{ref}} - s_{ij}$, where s_{ij} is the raw score subject i gave to stimulus j and s_{ij}^{ref} is the score given to the corresponding reference (original) video; then, the z-scores are computed: $z_{ij} = \frac{(d_{ij}-\mu_i)}{\sigma_i}$, where μ_i and σ_i are the mean and standard deviation of subject i's scores; These scores are then scaled to $[0,100]$: $z'_{ij} = 100\left(\frac{z_{ij}+3}{6}\right)$. The Differential Mean Opinion Score (DMOS) for each sequence is then computed: $\text{DMOS}_j = \frac{1}{N}\sum_{i=1}^{N} z'_{ij}$.Finally, to have a scale in which higher values means better quality, we compute the Reversed DMOS as: $\text{rDMOS}_j = 100 - \text{DMOS}_j$.

Results and discussion

Fig. 4a shows the histogram of all rates and Fig. 4b shows the MOS values for each source content. We can observe that, overall, when comparing the same content, the stereoscopic content is commonly rated worse than the monoscopic one. Such a behavior is also replicated in the distorted content, as will be discussed later. The "World tour" is an exception that can be explained by the fact that the pristine content itself is not perfect and also includes some stitching artifacts (see bottom part of Fig. 3(e)). Fig. 6 shows the processed Reversed DMOS values for each distorted sequence.

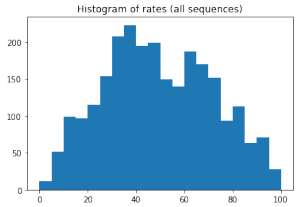
Objective metrics performance

For assessing the performance of objective metrics into the proposed subjective dataset, we use standard objective image quality metrics computed in the EAC *projection domain* and the same metrics *computed on viewports*, which better represents the final content as viewed by the user when wearing an HMD. The objective metrics computed on the projection domain are: Peak

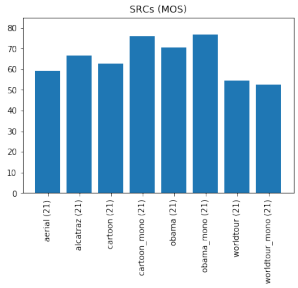


(a) Aerial (b) Alcatraz (c) Cartoon (d) Obama (e) World tour

Figure 3: Monoscopic versions of the source sequences used in the experiment



(a) Histogram of rates.

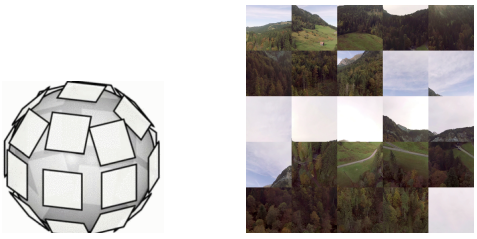


(b) Pristine content MOS.

Figure 4: Results based on the RAW scores.

signal-to-noise (PSNR); Structural Similarity (SSIM) [23]; Multi-scale Structural Similarity (MSSSIM) [22]; Visual Information Fidelity in pixel domain (VIFp) [16], and Gradient Magnitude Similarity Deviation (GMSD) [19].

The metrics above are also computed on the viewports, following the proposal of Birkbeck et al. [3]. For computing the viewport-based metrics, we pre-process the EAC video, generating N viewports for each frame, which are then merged into a collage frame (see Fig. 5). The collage frames are then used for computing the quality metrics on the viewports domain. Fig. 5 shows both the sampling pattern, named unifom, that we have used to compute the viewports and an example of a collage frame. (We have also tested the tropical and the equator sampling of [3], but the uniform sampling has provided the best correlation results due to a better coverage of the sphere.) In addition, for the configuration above, we have tested three different fields of view for the viewports: 30deg, 40deg, and 50deg.



(a) (b)

Figure 5: Example of uniform sampling (a) and collage frame (b) used to compute the viewport-based objective metrics.

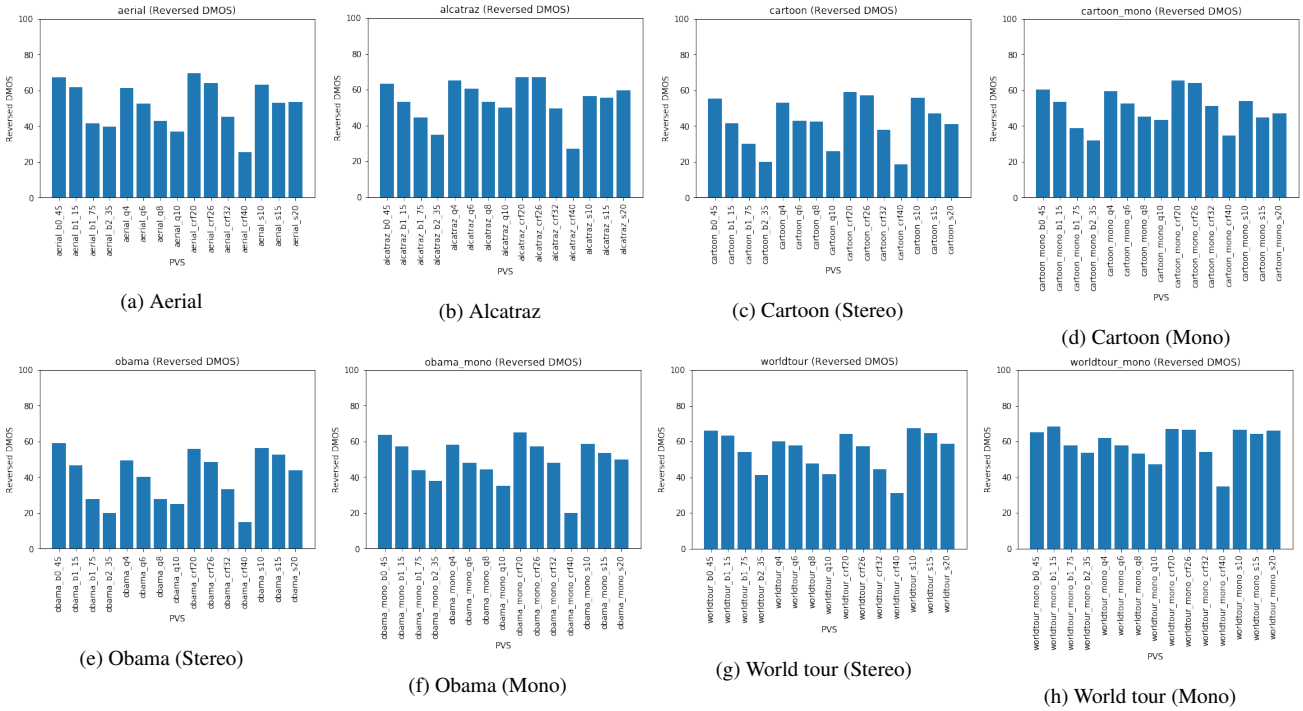
In both the projection domain and viewport-based versions, the objective metrics are computed individually for each frame and then pooled with an average method. Finally, the standard performance indexes Pearson linear correlation coefficient (PLCC) and Spearman rank order correlation coefficient (SROCC) are computed between the reversed DMOSs and the fitted logistic regression function: $s' = \frac{\beta_1 - \beta_2}{1 + e^{-\frac{s-\beta_3}{|\beta_4|}}} + \beta_2$.

Results and discussion

Fig. 7 shows the fitted logistic regression curve, and Table 1 shows the PLCC and SROCC results for the different metrics in our dataset. From the results, it is possible to conclude that: (i) the metrics that performed best on our dataset are SSIM for blockiness, MSSIM for H.264, and VIFP for blur and cubemap seams. (ii) overall, VIFP computed in the viewports using a 40-degree field of view is the best performing metric when considering all the distortions (correlation of 0.8421). (iii) when comparing the individual metrics in the different domains, there is a significant improvement when assessing the visual quality of the 360-degree content based on viewports instead of using the EAC projection domain. (It is expected that such a behavior is even more pronounced when the ERP format is used in the planar domain, which has more re-sampling and geometrical distortion than EAC.); (iv) the uniform sampling with 40-degree field-of-view provides the best performance for most of the metrics; (v) none of the metrics performs good enough on seams, which is a highly localized distortion that can be hidden on the global metrics; this result highlights the need for specific artifact metrics, such as [6]; and (vi) finally, it is possible to notice that the performance of the different metrics varies significantly according to the different distortion types and that there is not a single metric that performs the best for each individual distortion.

Conclusion

We established a subjective quality dataset for EAC 360-degree videos that includes specific visual distortion types and mixes both monoscopic and stereoscopic content. Then, we analyzed the performance of objective metrics computed both in the projection domain and in viewports extracted from the 360-degree content against the gathered subjective scores. Our experiments show that: visual distortions tend to affect more the quality of 360-degree stereoscopic content than its monoscopic counterpart; computing objective metrics in viewports provides a better correlation with subjective studies than computing them in the projection domain; and there is still a need for new perceptually-oriented metrics that can reliably measure specific visual distortions happening in processed 360-degree content, as it can be seen by the fact that there is not a single metric that performs the best for all the different distortions. As future work, we will extend the dataset to include other visual distortions that appear on 360-degree content [1], analyze the performance of stereoscopic metrics on our dataset, and design perceptually-optimized metrics for 360-degree videos that both consider different distortions and the user behavior when watching immersive content via HMDs.



(a) Aerial (b) Alcatraz (c) Cartoon (Stereo) (d) Cartoon (Mono)

(e) Obama (Stereo) (f) Obama (Mono) (g) World tour (Stereo) (h) World tour (Mono)

Figure 6: Reversed DMOS results.

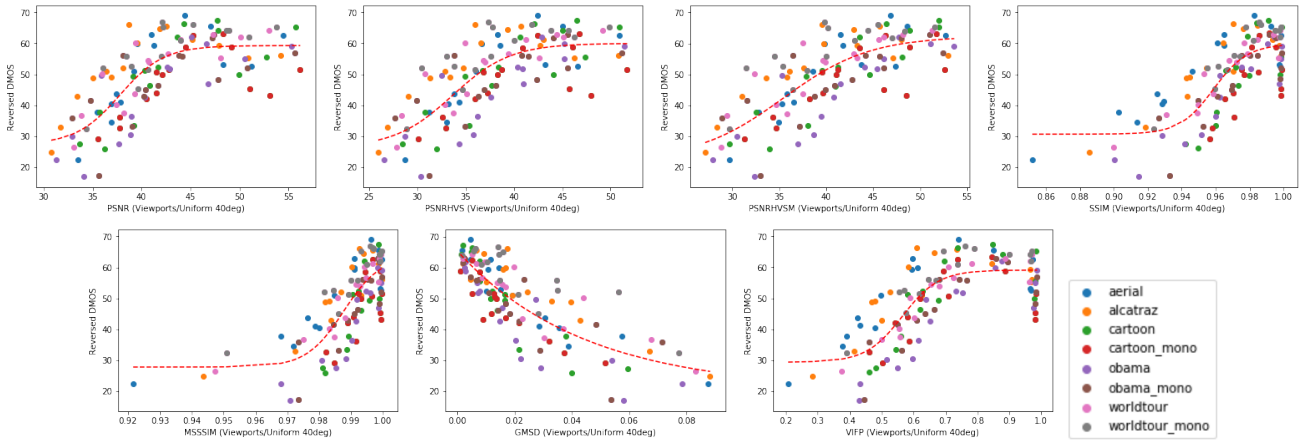


Figure 7: Frame-based objective metrics vs. subjective score plots with the fitted logistic function.

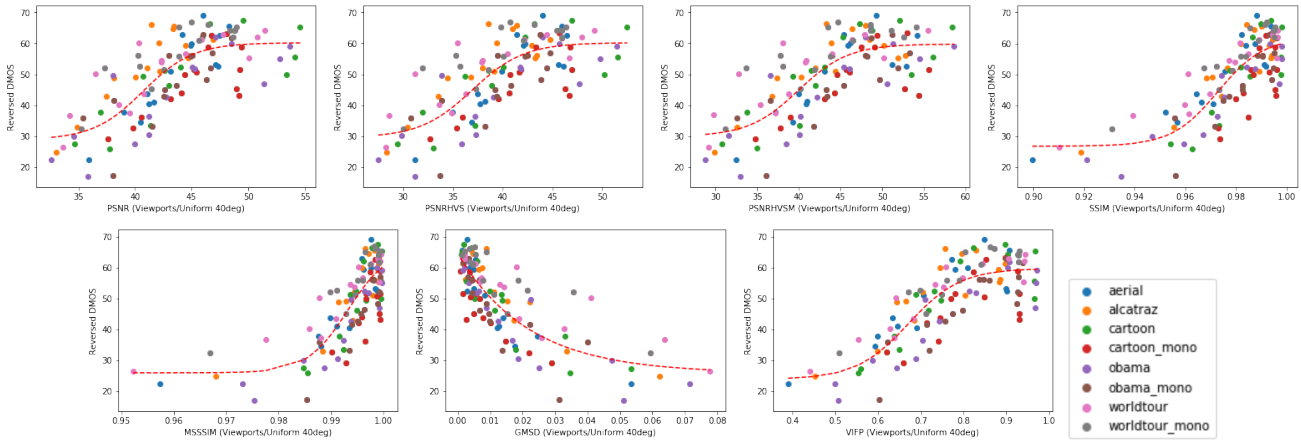


Figure 8: Viewport-based (uniform sampling with 40deg FoV) objective vs. subjective score plots with the fitted logistic function.

Table 1: Objective metrics performance (two best highlighted for each column).

Metric	Blur		Blockiness		Seams		H.264		All but seams		All	
Projection domain	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC
PSNR	**0.9102**	**0.85565**	0.6975	0.73387	0.37234	0.40391	0.87439	0.82588	0.80067	0.81114	0.7698	0.6984
PSNRHVS	0.90453	**0.85565**	0.72396	0.7544	**0.4752**	**0.49799**	0.87654	0.83761	0.79789	0.80956	0.77025	0.71917
PSNRHVSM	0.90127	**0.86783**	0.72272	0.75257	**0.4752**	0.39291	0.87206	0.83871	0.79951	0.8112	0.78126	**0.75778**
SSIM	0.8871	0.80261	**0.73684**	**0.76503**	-0.00633	0.01826	0.87597	0.83065	**0.83346**	**0.84026**	**0.79829**	0.70389
MSSSIM	0.87287	0.79043	0.71906	0.7423	-0.29563	-0.29826	**0.88409**	**0.84421**	**0.83293**	0.83507	0.79794	0.71481
GMSD	0.89075	0.84609	**0.76701**	**0.79362**	**0.4752**	**0.49799**	**0.89152**	**0.86217**	0.83136	**0.83626**	**0.82336**	**0.80654**
VIFP	**0.92313**	0.82522	0.62213	0.65946	**0.5464**	0.46704	0.87197	0.78959	0.82656	0.82954	0.79601	0.6891
Viewports domain (Mirage) / UniformFov30	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC
VP-U30-PSNR	**0.91901**	**0.87304**	0.66811	0.67485	0.3094	0.18103	0.86842	0.80425	0.81023	0.80901	0.78006	0.71927
VP-U30-PSNRHVS	0.9055	**0.86783**	0.70385	0.71554	0.28705	0.15712	0.88511	**0.84384**	0.81274	0.81835	0.78103	0.72291
VP-U30-PSNRHVSM	0.89778	0.86000	0.72625	0.73754	0.33252	0.30174	0.89208	**0.84384**	0.82147	0.82735	0.7866	0.71652
VP-U30-SSIM	0.89034	0.82000	**0.82190**	**0.82735**	0.29789	**0.30490**	0.89517	0.83138	**0.86308**	**0.84617**	0.83062	0.72872
VP-U30.MSSSIM	0.87516	0.82261	0.77593	0.7665	0.29491	0.21826	**0.90958**	0.84238	0.85678	0.83808	0.82564	0.72698
VP-U30-GMSD	0.88364	0.83478	**0.80219**	**0.80132**	**0.39352**	**0.38696**	0.90207	**0.85997**	0.84500	0.84335	**0.83167**	**0.80068**
VP-U30-VIFP	**0.92522**	0.86261	0.73905	0.75147	**0.34261**	0.27224	**0.90959**	0.83431	**0.87505**	**0.86176**	**0.84186**	**0.73738**
Viewports domain (Mirage) / UniformFov40	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC
VP-U40-PSNR	**0.89387**	0.82957	0.70128	0.69685	0.30829	0.08753	0.88145	0.82845	0.80472	0.80446	0.77457	0.70745
VP-U40-PSNRHVS	0.8796	**0.83304**	0.7226	0.72471	**0.3182**	0.07565	0.89239	0.84311	0.80397	0.81252	0.77197	0.70842
VP-U40-PSNRHVSM	0.87617	0.82957	0.73905	0.74927	0.31157	**0.21304**	0.89771	0.84384	0.81371	0.82339	0.77859	0.70633
VP-U40-SSIM	0.87759	0.81391	**0.83158**	**0.84128**	0.29789	0.3049	0.90306	0.83358	**0.86258**	**0.84947**	0.82861	0.72511
VP-U40-MSSSIM	0.86279	0.80783	0.7885	0.77419	0.25424	0.1513	**0.91662**	**0.85521**	0.85636	0.84201	0.82351	0.72004
VP-U40-GMSD	0.86537	0.81826	**0.8126**	**0.80938**	**0.40609**	**0.40261**	0.9081	**0.85997**	0.83936	0.83474	**0.82443**	**0.79005**
VP-U40-VIFP	**0.91709**	**0.86783**	0.74454	0.7489	0.11579	0.10438	**0.91598**	0.83138	**0.8764**	**0.86306**	**0.84214**	**0.72975**
Viewports domain (Mirage) / UniformFov50	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC
VP-U50-PSNR	0.8944	0.82435	0.73644	0.70931	**0.32978**	0.27834	0.89299	0.82735	0.8117	0.80921	0.78018	0.70505
VP-U50-PSNRHVS	**0.87971**	0.82609	0.73504	0.74707	**0.32978**	0.11142	0.89803	**0.85630**	0.80637	0.8139	0.77334	0.70028
VP-U50-PSNRHVSM	0.87642	**0.82522**	0.74008	0.76026	0.29801	0.16009	0.90048	0.85521	0.81314	0.82519	0.77747	0.70192
VP-U50-SSIM	0.8796	0.80609	**0.83011**	**0.84128**	-0.00867	-0.00087	0.90651	0.84787	**0.86528**	**0.85777**	**0.82952**	0.72452
VP-U50-MSSSIM	0.86259	0.81565	0.78856	0.77896	0.29789	**0.3049**	**0.91954**	0.85447	0.8586	0.84698	0.82371	0.71539
VP-U50-GMSD	0.86693	0.81826	**0.81726**	**0.81488**	**0.32185**	**0.3887**	0.91177	**0.86217**	0.84243	0.83899	0.81914	**0.77657**
VP-U50-VIFP	**0.91659**	**0.86087**	0.7389	0.7511	0.11579	0.09157	**0.91682**	0.83871	**0.87655**	**0.86715**	**0.84167**	**0.72665**

References

[1] R. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli, and P. Frossard. Visual distortions in 360-degree videos. *IEEE Trans. Circuits Syst. Video Technol.*, 2019.

[2] R. Azevedo, N. Birkbeck, I. Janatra, B. Adsumilli, and P. Frossard. On the first JND and break in presence of 360-degree content: An exploratory study. In *Proc. of the 11th MMVE*, MMVE'19, page 3, 2019.

[3] N. Birkbeck, C. Brown, and R. Suderman. Quantitative evaluation of omnidirectional video quality. In *Proc. 9th QoMEX*, pages 1–3, 2017.

[4] J. Boyce, E. Alshina, and Z. Deng. Subjective testing method for 360° Video projection formats using HEVC. Technical Report N16892, ISO/IEC, 2017.

[5] Z. Chen, Y. Li, and Y. Zhang. Recent advances in omnidirectional video coding for virtual reality: Projection and evaluation. *Signal Process.*, 146:66–78, 2018.

[6] F. De Simone, R. G. de A. Azevedo, S. Kim, and P. Frossard. Graph-based detection of seams in 360-degree images. In *IEEE ICIP*, pages 3776–3780, 2019.

[7] H. Duan, G. Zhai, X. Yang, D. Li, and W. Zhu. IVQAD 2017: An immersive video quality assessment database. In *IWSSIP*, pages 1–5. IEEE, 2017.

[8] M. Huang et al. Modeling the Perceptual Quality of Immersive Images Rendered on Head Mounted Displays: Resolution and Compression. *IEEE Trans. on Image Process.*, 27(12):6039–6050, 2018.

[9] ITU-T. P.910–subjective video quality assessment methods for multimedia applications. 2008.

[10] ITU-T. BT.500: Methodology for the subjective assessment of the quality of television pictures. 2012.

[11] J. Jerald. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, 2015.

[12] C. Li, M. Xu, X. Du, and Z. Wang. Bridge the Gap Between VQA and Human Behavior on Omnidirectional Video: A Large-Scale Dataset and a Deep Learning Model. In *Proc. of the 26th ACM MM*, Seoul, Republic of Korea, 2018.

[13] C. Li, M. Xu, S. Zhang, and P. L. Callet. State-of-the-art in 360deg video/image processing: Perception, assessment and compression. *arXiv preprint arXiv:1905.00161*, 2019.

[14] A.-F. Perrin, C. Bist, R. Cozot, and T. Ebrahimi. Measuring quality of omnidirectional high dynamic range content. page 38. SPIE, 2017.

[15] M. P. Queluz et al. Subjective and objective quality assessment of omnidirectional video. In *Applications of Digital Image Processing XLI*, page 25. SPIE, 2018.

[16] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Trans. Image Process.*, 15(2):430–444, 2006.

[17] E. Upenik, M. Řeřábek, and T. Ebrahimi. Testbed for subjective evaluation of omnidirectional visual content. In *PCS*, pages 1–5, 2016.

[18] M. Xu, C. Li, Y. Liu, X. Deng, and J. Lu. A subjective visual quality assessment method of panoramic videos. pages 517–522. IEEE, 2017.

[19] W. Xue et al. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. Image Process.*, 23(2):684–695, 2014.

[20] X. Yang, W. Ling, Z. Lu, et al. Just noticeable distortion model and its applications in video coding. *Signal Process.: Image Commun.*, 20(7):662–680, 2005.

[21] V. Zakharchenko, K. P. Choi, and J. H. Park. Quality metric for spherical panoramic video. *Proc. SPIE*, 9970:9970–9970–9, 2016.

[22] Zhou Wang et al. Multiscale structural similarity for image quality assessment. In *Proc. 37th IEEE Asilomar Conference on Signals, Systems and Computers*, 2003.

[23] Zhou Wang et al. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.

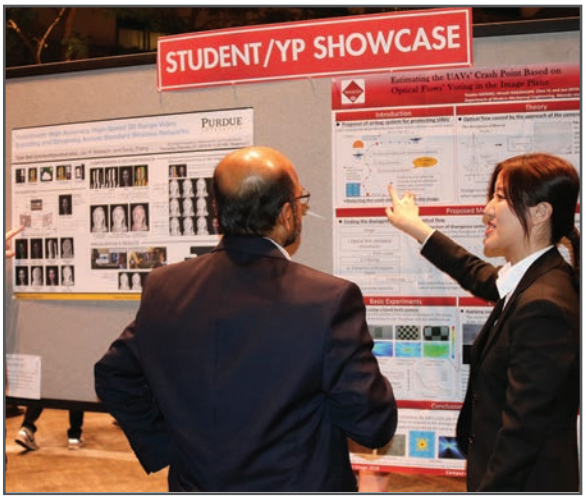
JOIN US AT THE NEXT EI!

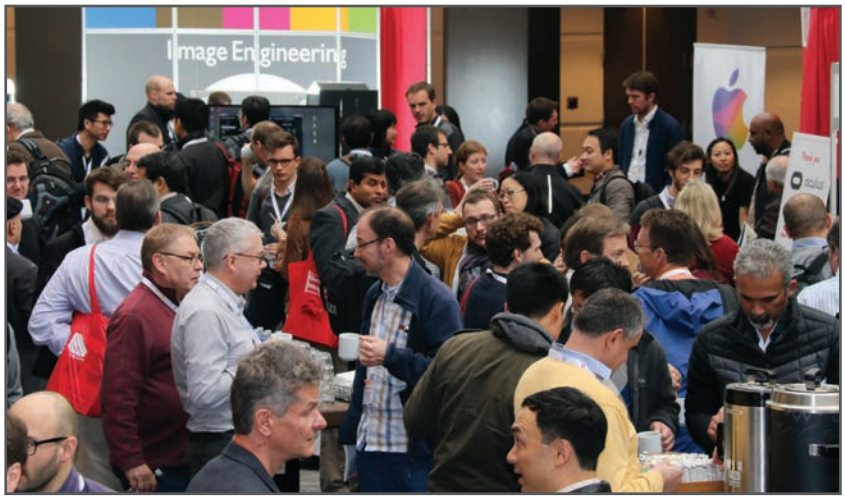
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org