On the Performance of Objective Metrics for Omnidirectional Visual Content

Evgeniy Upenik, Martin Rerabek and Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
Email: firstname.lastname@epfl.ch

Abstract—Omnidirectional image and video have gained popularity thanks to availability of capture and display devices for this type of content. Recent studies have assessed performance of objective metrics in predicting visual quality of omnidirectional content. These metrics, however, have not been rigorously validated by comparing their prediction results with ground-truth subjective scores. In this paper, we present a set of 360-degree images along with their subjective quality ratings. The set is composed of four contents represented in two geometric projections and compressed with three different codecs at four different bitrates. A range of objective quality metrics for each stimulus is then computed and compared to subjective scores. Statistical analysis is performed in order to assess performance of each objective quality metric in predicting subjective visual quality as perceived by human observers. Results show the estimated performance of the state-of-the-art objective metrics for omnidirectional visual content. Objective metrics specifically designed for 360-degree content do not outperform conventional methods designed for 2D images.

Keywords—*omnidirectional images and video; subjective quality evaluation; objective metrics; performance*

I. INTRODUCTION

Omnidirectional image and video, also referred to as 360-degree or cinematic virtual reality (VR) content, represent the form of immersive visual content providing viewers with an ability to look in all directions within a scene. The overall accessible field of view can cover a full sphere, a semi-sphere, or a spherical segment. Users are able to freely change the viewing direction by means of a head or device movement; it is also possible to navigate manually using gestures. Compression and visual quality assessment of omnidirectional images and video have recently attracted a lot of interest in scientific and engineering communities, mainly due to the latest technology developments simplifying the process of acquisition, delivery, and visualization of such 360-degree content.

Being intrinsically of a spherical nature, omnidirectional images and video are often represented in one of the geometrical panoramic projections suitable for their storage and visualization. Nowadays, two projections are widely used for VR content representation, namely, equirectangular and cubic. The first, an equirectangular projection, is well known as one of the geographical mappings [1]. The second, a cubic projection, originally comes from the field of computer graphics where it has been known for decades [2]. The latter, however, has many variations in omnidirectional imaging since cube faces can be ordered and placed differently in the mapped image frame.

360-degree images and video can be acquired with a single-sensor omnidirectional camera, a multi-sensor camera, or a set of cameras mounted on a rig. The latter two acquisition setups require an additional procedure referred to as a stitching process, which fuses pictures from different sensors together in order to form an omnidirectional image. Stitching is typically performed on-board in multi-sensor cameras and as a post-processing step in multi-camera systems. Single-sensor capturing devices, such as catadioptric [3], have become obsolete, whilst multi-camera[1] and multi-sensor[2] systems are now highly adopted by consumer electronics industry.

There are several ways to visualize and consume omnidirectional image and video content. The most immersive method requires subjects to wear a head-mounted display (HMD). A real-time head motion tracking of an HMD allows presentation of a part of the omnidirectional content (viewport) corresponding to the direction user looks at. A different way of consuming 360-degree images and video is via hand-held devices by exploiting their accelerometer features, which allows presentation of a desired viewport related to a pan-tilt position. Another less immersive approach is to visualize omnidirectional content on a regular screen assuming the navigation through the content manually, e.g. by using gestures.

Due to full-spherical coverage, omnidirectional visual content requires significantly higher resolution, when compared to conventional 2D content, to satisfy consumer expectations. Such requirements emphasize the importance of an appropriate compression algorithm to efficiently cope with delivery and storage of such a content. Evidently, existing image and video compression algorithms, developed for conventional 2D visual information, can be adapted to cope with 360-degree content. Nevertheless, a suitable compression algorithm must take into account the intrinsic spherical nature of omnidirectional content. Recent studies on compression attempt to improve its efficiency by addressing geometrical representations [4][5], adaptive delivery [6][7], and spherical nature of the visual content [8][9].

An important part of the development of future 360-degree image and video compression algorithms is related to the selection of objective metrics used to automate the process of visual

---

[1]http://vr.gopro.com
[2]http://theta360.com, http://ozo.nokia.com

(a) *Harbor*

(b) *Kite Flite*

(c) *Pole Vault*

(d) *Skateboard Trick*

Fig. 1: Test image contents used in the experiments.

quality assessment. So far, however, there is no agreement on which metrics should be used to predict perceived quality of omnidirectional content as there is not enough evidence about their performance. Recently, new objective metrics for omnidirectional content have been introduced [10][11]. Nonetheless, to the best of our knowledge, no attempt to benchmark their performance has been performed. To benchmark the available objective metrics, a ground-truth data is necessary. The most reliable way to obtain such ground-truth data is by means of subjective quality evaluation. Little to nothing has been done towards subjective evaluation of omnidirectional content. In [12], authors present new strategies for assessing the quality of composite video streams focusing on videoconferencing applications only. A testbed for subjective assessment of omnidirectional content using HMD is presented in [13] next to the results of a pilot subjective experiment.

The aim of this paper is to assess the performance of available objective metrics designed specifically for omnidirectional visual content against ground-truth subjective mean opinion scores (MOS). Additionally, a comparison to the performance of conventional 2D objective metrics has been carried out. The objective metrics are evaluated, in terms of commonly used performance indexes, i.e. linearity, monotonicity, accuracy, and consistency, based on their correlation with the perceived visual quality. It is shown that VIFP objective metric provide the best performance indexes. However, overall results indicate the need for new algorithms, which better predict perceived quality of omnidirectional content.

The remainder of this paper is structured in the following way. Section II describes a subjective evaluation experiment, including dataset, methodology, test environment, statistical analysis, and results. Section III presents a set of objective metrics used for benchmarking, their descriptions, performance evaluation, and discussion of results. Section IV concludes the paper.

Fig. 2: Cubic mapping with rotated faces packed as 3x2.

TABLE I: Spatial index (SI) and colorfulness (CF) computed for the test contents used in the experiment.

	Harbor	KiteFlite	PoleVault	SkateboardTrick
SI	7.96	10.45	10.33	6.31
CF	15.64	14.88	42.81	25.62

II. SUBJECTIVE EVALUATION

This section provides a detailed account of the experiment on subjective visual quality evaluation of omnidirectional content conducted by the authors. First, the dataset used for assessment is described and presented. Then the methodology and experimental environment are described. Finally, the analysis of results is presented followed by a discussion.

A. Dataset

Four high fidelity uncompressed omnidirectional images represented in equirectangular projection were used to produce an evaluation test-set; an additional image different from

TABLE II: Selected parameters and settings for all codecs exploited in the subjective experiments.

Codec	Software	Command line
JPEG	libjpeg	`jpeg -q quality referencefile.png compressedfile.jpg`
JPEG 2000	OpenJPEG	`opj_compress -q quality -i referencefile.ppm -o compressedfile.j2k`
HEVC	FFmpeg (x265)	`ffmpeg -f rawvideo -r 1 -s size -pix_fmt yuv420p -i infile.yuv -c:v hevc -crf qlt outfile.hevc`

those four was selected for training. The dataset is based on omnidirectional video test-set proposed by Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG. Omnidirectional video sequences were examined, and a still frame from each of the selected four sequences was taken to compose the dataset. Figure 1 depicts the selected contents in equirectangular projection. In order to keep the balance in spatial complexity and colors, spatial index (SI) [14] and colorfulness (CF) [15] were taken into account when selecting the images. Table I shows SI and CF for the contents used in the experiment.

Original images are represented in YUV color-space format with 4:2:0 chroma sub-sampling. Initial high resolution images were down-sampled using bi-cubic interpolation to a 3000×1500 pixels in order to correspond with the resolution of the HMD screen used in the experiments. Reference original images were then remapped to a cubic projection with rotated faces. This projection is a variation of a cubic mapping introducing the least amount of additional non-continuities (face edges) to the picture. Figure 2 shows an example of image represented in the cubic projection with rotated faces.

Both equirectangular and cubic images were compressed with three different codecs, namely JPEG, JPEG 2000, and HEVC intra. Afterwards, an expert screening was conducted to select bitrates representing the full scale of visual quality. As a result, four target bitrates, namely 0.25, 0.50, 0.75, and 1.00 bits per pixel, were selected. To compress original images with JPEG, JPEG 2000, and HEVC, the libjpeg[3], OpenJPEG[4], and FFmpeg with x265[5] software packages were used, respectively. The selected parameters and settings for all the codecs exploited in the subjective experiments are presented in Table II. In order to perform subjective assessments, all the encoded images were decompressed using the same respective software packages to produce reconstructed impaired stimuli.

B. Methodology

The experiment was conducted in the Multimedia Signal Processing Group (MMSPG) laboratory in EPFL where naïve subjects were invited to participate. It was performed according to Absolute Category Rating with Hidden Reference (ACR-HR) method described in [14]. ACR-HR is a single stimulus evaluation where the reference stimuli are randomly shown to observers among the impaired images. Stimuli are presented subsequently to subjects, and voting is performed after each viewing. Images are assessed using five-grade quality scale with the following levels: "5 - Excellent", "4 - Good", "3 - Fair", "2 - Poor", and "1 - Bad".

The observers were placed in an immersive environment where omnidirectional images were presented to them by means of an HMD. Immersive textual instructions were provided inside the VR along with a verbal guidance by the experimenter. Every test session started with an immersive training, consisted of three consequently shown images of a content not used in the evaluations. Subjects observed the examples of "Excellent", "Bad", and "Fair" quality levels shown in this particular order and were provided with the explanations and illustrations of impairment artifacts which can be found in the images. During the evaluation, subjects assessed the stimuli shown to them consequently without any time restrictions. When ready to rate an image subjects had to activate a 3D immersive voting menu by pressing a button and select the grade proceeding immediately to the next image. All stimuli were automatically randomized in each session.

All the steps described above in the current subsection including immersive training and evaluation were conducted using a testbed for subjective evaluation of omnidirectional visual content proposed in [13]. This software was developed for iOS and is publicly available for download[6]. It renders omnidirectional images with OpenGL using perspective projection and bi-cubic interpolation. The testbed allows uploading test stimuli to a device and changing immersive textual instructions. Voting data is acquired by the software and stored on the device. It can be further transmitted to a server for processing.

The following hardware equipment was used to perform immersive subjective quality evaluation of omnidirectional images along with the software testbed. During the experiment subjects were wearing an HMD composed of a VR head-mount with buttons[7] and a mobile device installed inside as a screen. iPhone 6 was used to display the images. The overall resolution of the phone's screen is 1334×750 pixels, which gives 667×750 pixels per eye. The vertical field of view provided by the hardware-software solution is 90 degrees and corresponds to 8.33 pixels per degree. All the subjects were sitting on a rotatable chair during the assessment.

Prior to the experiment a non-immersive training was provided to the subjects. The experimenter explained the purpose of the evaluation, showed examples of compression artifacts, and pointed to differences between coding and stitching artifacts. Subjects were instructed not to assess stitching artifacts. The test material, consisting of 104 test stimuli, was randomly distributed between two sessions. Each participant took part only in one session in order to shorten the time when the subject is exposed to VR immersive environment to a maximum of 25 minutes. Overall, 41 naïve subjects participated in the experiment. One subject was not able to complete the evaluations due to motion sickness. Subjects, 25
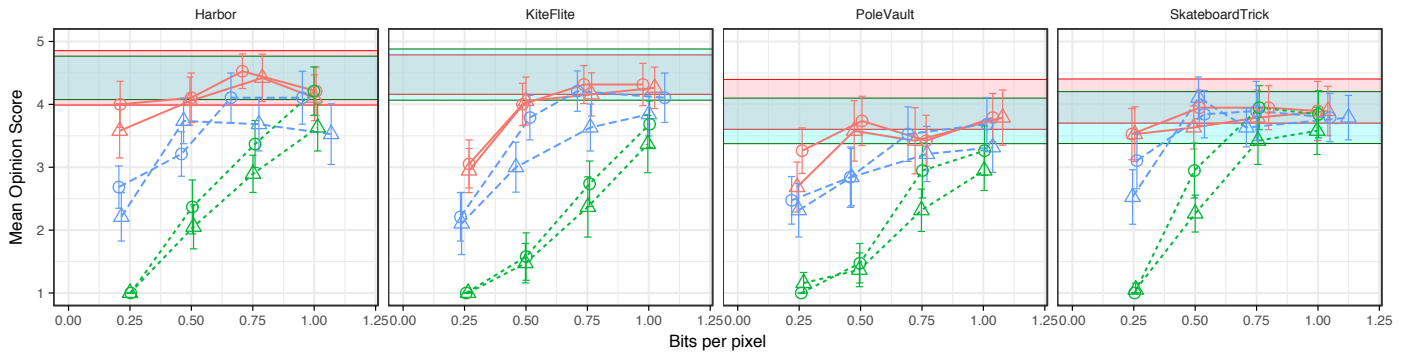
Fig. 3: MOSs with CIs obtained using ACR-HR method for compressed omnidirectional images. Red (solid) line represents HEVC encoded content, blue (long-dashed) - JPEG 2000, and green (short-dashed) - JPEG. Equirectangular projection is depicted with circles and cubic mapping - with triangles. Filled area between two horizontal lines corresponds to the 95% confidence interval of the hidden reference for each projection (red for equirectangular, cyan for cubic).

males and 15 females, were between 18 and 32 years old with an average and median of 24.9 and 24.8, respectively. All the participants were tested for correct color vision and visual acuity using Ishihara and Snellen charts respectively. An additional evaluation session was independently conducted with 5 expert viewers at 74th JPEG meeting in Geneva.

C. Data Analysis

Outliers detection was performed separately on the raw experimental data from each of two test sessions, since an individual subject had only assessed stimuli from one subset. Boxplot based method was used to remove outliers in the same way as in [16]. One subject was detected as an outlier in the first session. Therefore, to preserve the symmetry of the data, one randomly selected subject was removed from the second session. MOS values were computed for each stimulus in the entire dataset as mean values for the set of scores provided by different subjects. To estimate statistical significance, 95% confidence intervals (CI), assuming a Students t-distribution of the scores, were computed alongside with MOS values.

D. Results and Discussion

Figure 3 shows MOSs and CIs plotted for different contents. MOSs obtained from naïve subjects were shown to be highly correlated with expert subjects results. More particularly, standard correlation indexes between naïve and expert scores are *PLCC = 0.95, SROCC = 0.87, RMSE = 0.40*. Certainly, this allows us to consider the subjective evaluation results being reliable and consistent.

The results of the subjective evaluation experiment and the data analysis show, as expected, higher performance of HEVC and JPEG 2000 when compared to JPEG at lower bitrates. Some of the contents, however, namely "Pole Vault" and "Skateboard Trick", are systematically underrated, which can be possibly explained by the lower perceptual quality of the original pictures. Other explanations for the former can be the following. There are many human faces in the "Pole Vault" content, and thus, due to a relatively low resolution of the HMD screen, observers' expectations to distinguish facial

features were not met. In the "Skateboard Trick" content, there is an artificially blurred circle below the camera used to camouflage a tripod, which could influence the decision of naïve subjects.This hypothesis is supported by the fact that in the expert subjective results "Skateboard Trick" reference stimuli was not underrated, whilst "Pole Vault" was.

When compressing images represented in a cubic projection, edges of continuous parts of the frame are distorted non-uniformly with different intensity. This makes cube-face borders distinguishable for some stimuli in the rendered viewport when observed using an HMD. Experimental results, indeed, show lower scores for cubic mapping at medium bitrates and the same scores as for equirectangular mapping at high and low bitrates. This may occur for the reason that at high bitrates there are no impairments, and at low bitrates the entire image is distorted, thus only at medium bitrates the cube-face borders are distinguishable due to compression artifacts.

III. OBJECTIVE EVALUATION

This section presents objective evaluation data for omni-directional visual content obtained by calculating particular metrics. Performance of these metrics is then evaluated by comparison to the ground-truth subjective scores. Finally, the results are presented alongside with a discussion.

Omnidirectional visual content can be assessed with conventional 2D objective metrics as well as with metrics designed specifically for 360-degree images. Here we provide a list of objective evaluation methods used in this study. The following objective metrics were computed:

- Conventional 2D metrics
  1) *Peak signal-to-noise ratio (PSNR)*
  2) *Structural Similarity (SSIM)*
  3) *Multi-Scale Structural Similarity (MSSSIM)*
  4) *Visual Information Fidelity in pixel domain (VIFp)*
- Metrics designed for omnidirectional visual content
  1) *Spherical PSNR (S-PSNR)* computes PSNR for the set of points uniformly distributed on a spherical surface, where
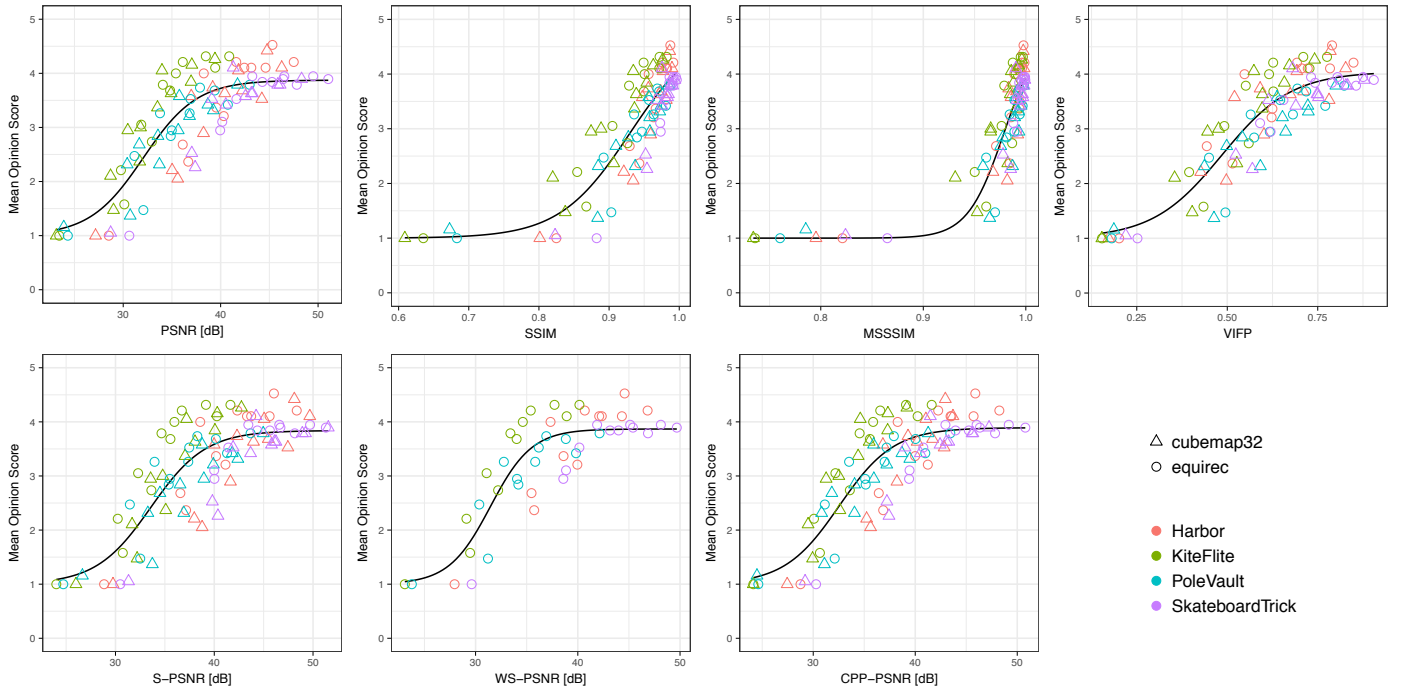
Fig. 4: Mapping of objective scores to subjective ratings. Triangles represent cubic projection, circles represent equirectangular projection. Different contents are marked with colors: "Harbor" - red, "Kite Flite" - green, "Pole Vault" - cyan, "Skateboard Trick" - magenta. Solid black line depicts a logistic fitting.

corresponding pixels from a reference and an assessed image are reprojected to this set [10].

2) *Weighted Spherical PSNR (WS-PSNR)* computes PSNR in such a way that intermediate values for pixels in equirectangular image of height h are weighted with a coefficient $w_{i,j} = cos((i - h/2)\pi/h)$ [17]. This weighting reduces the impact of the pixels with higher latitudes. It should be noted that WS-PSNR is only applicable for images in equirectangular projection.

3) *Craster Parabolic Projection PSNR (CPP-PSNR)*. Both an assessed image and a reference are re-mapped to a Craster parabolic projection, then PSNR is computed in that domain [11].

To compute conventional objective metrics, namely PSNR, SSIM, MSSSIM, and VIFP, a publicly available software package *VQMT*[8] was used. For metrics designed specifically for omnidirectional content, S-PSNR, WS-PSNR, and CPP-PSNR, publicly available *Samsung 360 Tools*[9] were used.

A. Performance Evaluation

Standard performance indexes, namely, the Pearson linear correlation coefficient (PLCC), the Spearman rank order correlation coefficient (SROCC), the Root mean square error (RMSE), and the Outlier ratio (OR), were computed to compare objective results with the ground-truth subjective ratings. To calculate the above listed performance coefficients, the raw objective evaluation data was fitted to the MOS values.

Logistic fitting was performed considering that the data is in different scope and in order to compensate possible saturation of subjective scores. One can see the fitted curves in Figure 4.

Table III presents linearity, monotonicity, accuracy, and consistency indexes. These indexes were computed assuming different mapping schemes of test data:

- *A* - for equirectangular projection, on all the contents,
- *B* - for cubic projection, on all the contents,
- *C* - for both projections, on all the contents,
- *D* - for both projections, each content separately.

More specifically, for A and B cases the fitting was performed only for the data points representing each individual projection, for C and D cases the fitting was performed on all the contents to compute the indexes. Moreover, for D case, an average of resulted indexes for each content was considered.

B. Results and Discussion

Figure 4 shows scatter plots of MOS values against objective metrics. For the cases A, B, and C, the results of the objective metrics performance evaluation show only moderate correlations with the ground-truth subjective scores and do not significantly change for different projections. As it can be seen from scatter plots in the Figure 4, points are sparse and not concentrated along the fitting curve. Moreover, objective metrics designed specifically for omnidirectional visual content do not show better performance when compared to common objective quality evaluation measures. For the case D, performance per content is significantly higher compared to cases A, B, and C.

TABLE III: Standard performance indexes. Subcolumns A, B, and C, represent the results for equirectangular, cubic, and both projections computed over all the contents, respectively. Subcolumn D shows an average of coefficients computed for each content separately.

Metric	PLCC				SROCC				RMSE				OR			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
PSNR	0.8714	0.8437	0.8553	0.9487	0.7176	0.7731	0.7567	**0.8909**	0.4804	0.5103	0.5008	0.2929	0.4375	0.4375	0.4167	**0.2396**
SSIM	0.8898	0.8632	0.8740	0.9459	0.7365	0.7927	0.7709	0.8821	0.4464	0.4790	0.4689	0.3050	**0.3958**	0.4583	0.4167	0.2812
MSSSIM	0.9059	0.8661	0.8860	0.9123	0.7539	0.7796	0.7814	0.8394	0.4143	0.4755	0.4483	0.3887	0.4583	0.4167	0.4271	0.3229
VIFP	**0.9116**	**0.8875**	**0.8994**	0.9319	**0.7608**	**0.8029**	**0.7953**	0.8538	**0.4025**	**0.4374**	**0.4221**	0.3395	**0.3958**	**0.3958**	0.4167	0.3125
S-PSNR	0.8766	0.8482	0.8392	0.9168	0.7376	0.7836	0.7307	0.8214	0.4715	0.5035	0.5257	0.3705	0.4583	0.4375	0.4271	0.3021
WS-PSNR	0.8748	-	-	**0.9583**	0.7297	-	-	0.8648	0.4746	-	-	**0.2544**	0.4375	-	-	0.2500
CPP-PSNR	0.8800	0.8521	0.8658	0.9467	0.7403	0.7745	0.7697	0.8843	0.4654	0.4975	0.4838	0.2966	0.4375	0.4167	**0.4062**	0.2500

However, conventional metrics still outperform those designed for 360-degree content.

Since S-PSNR, WS-PSNR, and CPP-PSNR are all based on PSNR, it is reasonable to compare them mutually. Looking at the scatter plots in the Figure 4, one can notice that the distribution patterns of the score points are of high similarity for all the PSNR based metrics showing strong content dependency.

IV. CONCLUSION

In this paper we have provided the results of a subjective evaluation experiment on omnidirectional images. A total number of 45 observers were involved in the study, including 40 naïve and 5 expert participants. Subjective evaluation scores were obtained for 104 test stimuli. Seven objective metrics, among which three are specifically designed to assess omnidirectional visual content, were calculated for each stimuli. Rigorous performance evaluation has been carried out for objective quality metrics for omnidirectional visual content.

Analysis of the obtained subjective and objective scores indicates moderate performance of investigated metrics for omnidirectional visual content. Being PSNR based, these metrics do not outperform significantly their ancestor in predicting visual quality of omnidirectional content. All the evidence above suggests that the problem of better objective quality evaluation methods for omnidirectional visual content remains open. The future work should consider developing a more suitable objective metric for 360-degree content.

ACKNOWLEDGMENT

REFERENCES

[1] J. P. Snyder, *Flattening the Earth: Two Thousand Years of Map Projections*. University Of Chicago Press, 1993.

[2] N. Greene, "Environment mapping and other applications of world projections," *IEEE Computer Graphics and Applications*, vol. 6, no. 11, pp. 21–29, 1986.

[3] S. Nayar, "Catadioptric omnidirectional camera," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Proceedings*, 1997, pp. 482–488.

[4] I. Bauermann, M. Mielke, and E. Steinbach, "H.264 Based Coding of Omnidirectional Video," in *Computer Vision and Graphics*, ser. Computational Imaging and Vision, K. Wojciechowski, B. Smolka, H. Palus, R. S. Kozera, W. Skarbek, and L. Noakes, Eds. Springer Netherlands, 2006, no. 32, pp. 209–215.

[5] J. Li, Z. Wen, S. Li, Y. Zhao, B. Guo, and J. Wen, "Novel tile segmentation scheme for omnidirectional video," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 370–374.

[6] P. Rondao Alface, J.-F. Macq, and N. Verzijp, "Interactive Omnidirectional Video Delivery: A Bandwidth-Effective Approach," *Bell Labs Technical Journal*, vol. 16, no. 4, pp. 135–147, Mar. 2012.

[7] M. Hosseini and V. Swaminathan, "Adaptive 360 VR Video Streaming: Divide and Conquer!" *arXiv:1609.08729 [cs]*, Sep. 2016. [Online]. Available: http://arxiv.org/abs/1609.08729

[8] I. Tosic and P. Frossard, "Low bit-rate compression of omnidirectional images," in *Picture Coding Symposium, 2009. PCS 2009*, May 2009, pp. 1–4.

[9] F. De Simone, P. Wilkins, N. Birkbeck, A. Kokaram, and P. Frossard, "Geometry-driven quantization for omnidirectional image coding," in *32nd Picture Coding Symposium (PCS 2016)*, 2016.

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

[11] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," vol. 9970, 2016, pp. 99 700C–99 700C–9.

[12] S. Leorin, L. Lucchese, and R. G. Cutler, "Quality assessment of panorama video for videoconferencing applications," in *IEEE 7th Workshop on Multimedia Signal Processing*, 2005, pp. 1–4.

[13] E. Upenik, M. Rerabek, and T. Ebrahimi, "A testbed for subjective evaluation of omnidirectional visual content," in *32nd Picture Coding Symposium PCS*, Dec 2016.

[14] ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, April 2008.

[15] D. Hasler and S. E. Suesstrunk, "Measuring colorfulness in natural images," vol. 5007, 2003, pp. 87–95.

[16] F. De Simone, L. Goldmann, J.-S. Lee, and T. Ebrahimi, "Towards high efficiency video coding: Subjective evaluation of potential coding technologies," *Journal of Visual Communication and Image Representation*, vol. 22, no. 8, pp. 734–748, Nov. 2011.

[17] Y. Sun, A. Lu, and L. Yu, "WS-PSNR for 360 video quality evaluation," Proposal, MPEG2016/M38551, ISO/IEC JTC1/SC29/WG11, Geneva, May 2016.