IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 13, 2020, accepted February 10, 2020, date of publication February 14, 2020, date of current version February 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974010

Subjective Assessment of 360° Image Projection Formats

TRAN THI HAI UYEN, OH-JIN KWON, SEUNGCHEOL CHOI, AND IKRAM HUSSAIN
Department of Electrical Engineering, Sejong University, Seoul 05006, South Korea

Corresponding author: Oh-Jin Kwon (ojkwon@sejong.ac.kr)

This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) Grant funded by the Korea Government (MSIT) (No. 2017-0-01667, Development of the next-generation 360° image/video format).

ABSTRACT Recently, the use of virtual reality has been spreading rapidly in many industrial fields, such as games and entertainment, and 360° contents are widely used in various applications of the virtual reality (VR) market. The 360° contents create a virtual space comparable to real scenes. Coupled with the rapid growth of the 360° image/video field, many projection formats have been developed and published. However, the studies which identify the best format for a variety of applications are still limited. Therefore, it is desirable to evaluate the visual quality of the projection formats used in the current market, and the result of the evaluation can be used as an important consideration capable of providing 360° VR service of better quality. Here, we propose a methodology for the subjective test and briefly present ten 360° projection formats used in our experiment. The proposed methodology for the subjective test uses a combination of the Degradation Category Rating (DCR) and Pair Comparison methods. We performed the subjective test with 15 participants for each test step, the experimental set-up for which followed the proposed methodology. The results show that the equirectangular projection format (best known and widely used format in most commercial products) did not give the highest quality. Hybrid equiangular cubemap and equatorial cylindrical projection gave the best quality compared with the other eight formats. Based on the results of the DCR, it is noted that different monitors (size and resolution) do not affect the ranking of 360° image projection formats. Our work presents the ranking of ten published projection formats which is useful for identifying the best format for a future 360° image standard.

INDEX TERMS 360° projection format, subjective evaluation, degradation category rating, pair comparison, 360° image.

I. INTRODUCTION

Currently, with the rapid advance of virtual reality (VR), which provides a ''real life'' and ''being there'' experience, the 360° image is widely used in many applications in the VR industry [1]–[3]. The 360° image contains an omnidirectional scene in the 360° horizontal and 180° vertical direction such that it can provide a rich experience to the user, whereas a conventional image contains a scene of fixed field of view (FoV) [4], [5]. The 360° image normally allows the user to freely change his or her viewpoint and dynamically view other scenes. Since one viewport at any one time is normally shown to the user in 360° image services, it is emphasized that the quality of the viewport displayed to the user determines the overall quality of the 360° image.

The associate editor coordinating the review of this manuscript and approving it for publication was Simone Bianco.

In most 360° image services, the two-dimensional (2D) image is generated by unfolding the pixel information from the spherical space of the 360° image to the 2D plane [6]–[8]. This operation is called projection and the main reason for this is to use already existing high efficient 2D image coding techniques because no efficient encoding techniques have been developed for the images represented in the spherical space so far. Various projection methods have been introduced in the industry. Equirectangular projection (ERP) is the most widely used format in the industry because it is intuitive and easy to generate. For example, Google's Street-view provides a service that allows the user to navigate a location in all directions using an ERP image.

Based on the increasing interest in 360° image and video, development of an international standard for creating, exchanging, and storing 360° content has begun, and a draft version of the standards are being prepared.

From the standpoint of the standardization, the direction of the standardization for 360° image and video is roughly divided into two aspects: the coding efficiency and the metadata, which defines the format of applications that provide services using 360° image and video.

From the coding efficiency point of view, activities for exploring an optimal projection in terms of image and video coding are continuously conducted by JPEG (ISO/IEC JTC1 SC29 WG1) and MPEG (ISO/IEC JTC1 SC29 WG11), respectively. In the JPEG case, even though an investigation on various projection methods is being undertaken, an image coding standard for the 360° image has not yet been established. However, the MPEG began the standardization of the next generation video compression standard called Versatile Video Coding (VVC), including compression tools for 360° video. VVC will support projection formats that show better coding performance and additional features that enhance the performance of the coding, such as packing method, etc. [9]

Meanwhile, both JPEG and MPEG are developing international standards to define metadata describing the 360° image and video based on the ERP format: JPEG 360 Metadata [10] and the Omnidirectional Media Format [11], respectively. They specify the metadata that is required to store, exchange, and play the 360° image and video.

Although both standardizations are proceeding on different aspects, the definition of the projection format is common to both. The projection format that can be efficiently used in various applications is designated as the default format, and the metadata describing the 360° contents and their behavior are specified in both standards. In addition, the projection format for generating 360° image and video is regarded as a key element of a VR service using 360° image and video.

Therefore, analysis and quality assessment of the various projection formats that are currently used in the market is a useful and a beneficial research topic for an industry that needs to provide 360° content effectively and reliably. Furthermore, as described earlier, the quality of the viewport should be considered crucial in this analysis and quality assessment because the 360° scene is always displayed to the user by generating the viewport, and the user watches the 360° scene by changing the viewport. The viewport of each projection format is interpolated differently corresponding to the characteristics of the projection format. From this perspective, it is noteworthy that a projection format that provides the best quality viewport can provide a better quality of service using 360° image and video.

In this study, we evaluate the visual quality of 360° projection formats by comparing the viewport image subjectively, whereas other studies performed their evaluations by comparing the coding efficiency of the 360° projection format. We generate sample images using different projection formats from the high quality ERP image and subjectively evaluated their quality of the viewport image. We aim to select the most appropriate projection format that could be used in many applications as a common format. To achieve this goal, we propose a framework for the subjective evaluation of 360° image projection formats, and to the best of our knowledge, we uniquely conduct a subjective evaluation of the 360° projection formats.

The rest of the paper is organized as follows. Related work, including a brief overview of the ten 360° projection formats and the subjective test methods, is presented in Section II. The proposed subjective test method for the 360° projection formats is presented in Section III. Section IV presents the experimental results. The analysis of the results is presented in Section V, while the conclusions are given in Section VI.

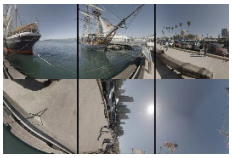
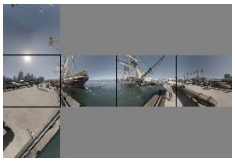
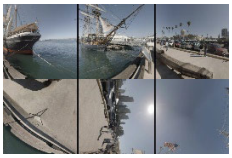
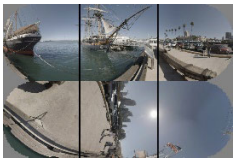
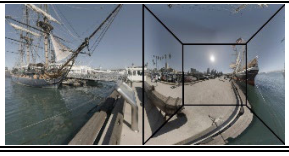
II. RELATED WORK

A. JOINT VIDEO EXPLORATION TEAM (JVET)

The JVET team was established in October 2015 by ITU-T SG16/Q6 VCEG and MPEG for exploration of future video coding technology. A part of the JVET's work is to explore technologies related to 360° video that can be included as a tool in the future video coding standard. They have explored several 360° projection formats to propose a format that shows the best coding performance requisite for a future standard. Furthermore, they have developed a 360Lib software package [12] for 360° video coding and processing. The software supports the ten 360° projection formats [13]: ERP, adjusted equal-area projection (AEP), cubemap projection (CMP), octahedron projection (OHP), truncated square pyramid projection (TSP), adjusted cubemap (ACP), rotated sphere projection (RSP), equatorial cylindrical projection (ECP), equiangular cubemap (EAC), and hybrid equiangular cubemap (HEC). The 360Lib software is used as an important tool for conversion between the projection formats, especially when other projection formats are converted from/to the ERP format. Detailed features of the ten projection formats tested herein are presented in Table 1. The 360° image is unlike a conventional 2D image. It is impossible to present a whole 360° in the image at one time as in a normal 2D image. Instead, a viewport presenting part of the image is rendered and displayed directly to the participant by normal displays. The way that objects in the viewport of a 360° image are presented to the user is dependent upon its spatial location with respect to the point of view (PoV) and FoV that are decided by the user.

The 360Lib supports the generation of the viewport for several projection formats. The viewport is generated by conversion from a 3D to a 2D object by rectilinear projection [11] along with the given PoV and FoV information. FoV defines the area displayed in the viewport illustrated in Fig. 1. Fig. 2 shows the rendered viewports in different FoVs. When the FoV is large, a wide scene is shown in the viewport. Likewise, when the FoV is small, a narrow scene is shown in the viewport.

TABLE 1. Brief introduction to ten projection formats.

Format	Features	2D sample	References
ACP	- Adjusted format from CMP by modifying the look-up vectors. It enlarges the center of the cubemap.		[7], [14]
AEP	- Modified version of equal-area projection format. AEP stretches the sphere vertically to get closer to the north and south poles.		[7], [14], [15]
CMP	- CMP deforms a sphere into a cube, then unfolds the cube to six faces.		[7], [14]
EAC	- EAC aims at having a uniform sampling rate regardless of sampling location on a cube face. EAC maintains a sample of equal lengths and creates a uniformly allocated pixel.		[7], [14], [16]
ECP	- ECP projects the equatorial region of the sphere using the Lambert cylindrical equal-area projection. It also projects the poles of the sphere onto square faces to avoid inactive pixels.		[17]
ERP	- The most widely used format for representing 360° images. ERP maps meridians (latitudes) and circles (longitudes) to the vertical and horizontal axes, respectively, by spacing them equally on the 2D plane. The north pole and south pole are stretched across the entire upper and lower edges of the flattened grid, respectively.		[7], [14]
HEC	- HEC is modified from CMP with the purpose of improving the sampling distribution. It applies an equal-angular mapping formula for all six faces.		[14]
OHP	- The 3D map of OHP is an octahedron. It contains eight triangular faces. The original 2D projection format contains non-active pixel areas. An improved version of OHP has two packing types without the non-active pixel areas.		[7], [14]
RSP	- RSP segments the sphere into two parts: bottom and top segments which are arranged like a tennis ball. It can be visualized as having six faces. The segments are packed similarly to ERP.		[7], [14]
TSP	- TSP utilizes cubic geometry and wraps the six cube faces into a compact frame. The front face of TSP corresponds to the front face of CMP. The back face of TSP is subsampled by four in both horizontal and vertical directions, while the side of the TSP format is created by warping the side of the cube faces to the trapezoidal region.		[14], [18]

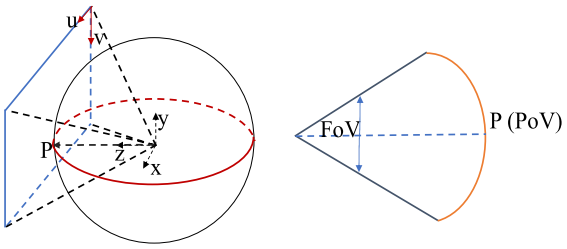


FIGURE 1. Illustration of FoV and PoV.

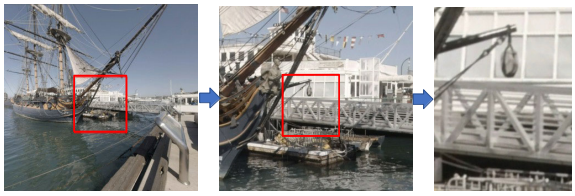


FIGURE 2. Viewport (red square) in different FoVs. (a) FoV = 90°, (b) FoV = 30°, (c) FoV = 10°.

B. SUBJECTIVE TEST METHODS

Subjective test methods have been widely used in the field of image quality evaluation. The subjective test is a psychophysical experiment that obtains a result by analyzing the means of subjects' judgment data. It is the most reliable method of image quality measurement. The specific design of the test is dependent upon the purposes of the test. Thanks to its reliability, the results of the subjective evaluations can be used as the ground truth data for developing new algorithms such as image coding tools or objective evaluation algorithms [19], [20]. However, there are limitations to this method, as the subjective test is extremely time-consuming and requires a huge human resources operation [21].

Numerous methods have been investigated to improve the performance and accuracy of the subjective test. Based on the requirements of the experiment, it was necessary to carefully select the most effective method. In the International Telecommunication Union (ITU-T P.910) publication, several subjective test methods for video quality assessment were defined [22]. In the case of a quality test with reference, the Degradation Category Rating (DCR) method is recommended [22]. In addition, the Pair Comparison (PC) method is recommended for highly discriminatory results. The PC method is well-suited for test materials that are nearly equal in quality. Furthermore, a combination of DCR and PC is recommended when there are many test cases [22]. This combination is useful for reducing the number of test cases while still giving high accuracy results.

In the DCR method, the test images are presented with a reference source at the same time. After each presentation, the participants evaluate the quality of the test image shown. Normally, there is a five-level scale for rating the overall quality of the test images, where 5 represents imperceptible and 1 represents very annoying. A limited time is normally given for their evaluation of the overall image quality. The DCR method's stability and its discriminatory ability are superior to other comparable methods [23].

PC is a process of comparing images in a pair. All possible combinations of two images should be compared. Participants decide which image shows better quality. In all cases of PC, the participant is forced to choose one image, even if the quality difference is hardly noticeable. The time available to make choices for this method is unlimited.

Mantiuk *et al.* [19] performed experiments on a comparison of four subjective methods for the image quality assessment. They evaluated four methods: single stimulus, double stimulus, forced choice pairwise, and similarity judgments in their experiments and concluded that the forced choice pairwise method gave the most accurate and time-efficient results. However, in the case where many items need evaluation, the PC test would generate too many test cases when generating the pairs of test items [22]. Therefore, a method to reduce the number of test cases before performing the PC test was required. Tran Thi Hai *et al.* [24] evaluated the quality of the 360° images using an absolute category rating method. They reported that the ERP scored lower than others in subjective image quality.

Other researches on 360° video and image have focused on the coding performance that use head-mounted displays (HMD) to perform the subjective test [5], [6], [25]. However, the HMD does not support the new projection formats. In our research, we generated the same viewport as the viewport shown in HMD.

III. PROPOSED METHOD

Owing to the large number of projection formats that have been published, users and researchers are confused when they wish to select a good quality format for their works. Furthermore, evidence must be produced to allow selection of a high-quality projection format for a future 360° image standard. Therefore, it is necessary to have an experiment to evaluate the quality of well-known projection formats. For these reasons, we propose a framework for the subjective test of the 360° projection formats via viewport images. We consider the ten projection formats: ACP, AEP, CMP, EAC, ECP, ERP, OHP, TSP, RSP, and HEC as listed in Table 1.

The proposed framework for the subjective test of the 360° image projection formats is shown in Fig. 3. Currently, the ERP format is the most used and supported by 360° camera vendors as the original format when they stitch the images taken from several lenses of the camera. For this reason, we use the ERP format as the original format to generate other formats. The 4K resolution images of the ten projection formats were generated by down sampling the original 8K (8192 × 4096) resolution ERP image using 360Lib. In addition, the number of active pixels to display in the viewport needs to be the same for all ten projection formats. In case of CMP and OHP, they have non-active pixels those are outside

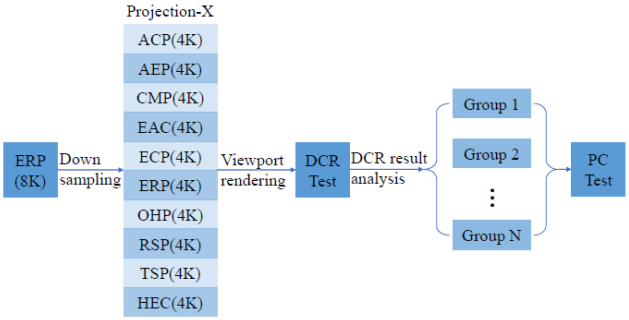


FIGURE 3. Framework for the evaluation of the ten 360° projection formats.

TABLE 2. Resolution Of projection formats.

ACP	AEP	CMP	EAC	ECP
3324 × 2216	3840 × 1920	4432 × 3324	3324 × 2216	3324 × 2216
ERP	**OHP**	**RSP**	**TSP**	**HEC**
3840 × 1920	5424 × 2712	3324 × 2216	3840 × 1920	3324 × 2216

of the faces, dark gray pixels, in the images as shown in Table 1. Depending on the properties of the arranged active pixels, the 4K image size of each format was decided as shown in Table 2. (It is not possible to set all the projection resolutions be the same because the shape (the ratio of width to height) of arranged active pixels of a projection is different from another. We set the image resolution be the closest to 4K for each projection format with maintaining the shape of the projection.)

The PC test requires 45 combinations for the ten projection formats totally. This leads too long and fatiguing works when participants evaluate each projection format through several viewports. To solve this problem, as recommended in [22] and [26], we propose a combination of DCR and PC for our subjective test to reduce the number of test cases. We first undertake the DCR test in our framework as shown in Fig. 3. After all of the participants finished the DCR test, those projection formats are divided into several groups, and subjected to the PC test. A group is a set of projection formats showing a relatively less noticeable visual difference in the DCR test. The PC test is performed independently for each group.

We have implemented our evaluation software based on the specified procedure of the DCR and PC test. To reduce the effects of reflected illumination, gray color is chosen as the background of the test screen, as recommended by the ISO [27].

The framework for the DCR test is presented in Fig. 4. The 8K ERP image is used as a ground truth image. A viewport, which is unfolded from the 4K projection image (Projection X), is compared to the viewport from the ground truth image. A viewport of the ground truth image and a corresponding viewport of the Projection X are displayed simultaneously, and the participant rates the quality of the Projection X's viewport compared to the ground truth's viewport.

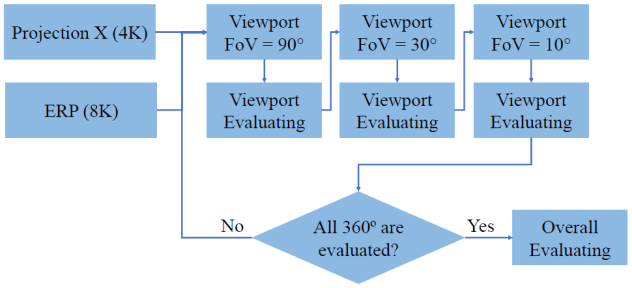


FIGURE 4. The proposed DCR test framework.

Three FoVs: 90°, 30°, and 10° are used for the evaluation of Projection X. In order to cover the entire sphere fully, we use six viewports at 90°, as shown in Fig. 5, which have different PoVs from V1 to V6, as shown in Fig. 6, respectively. The 30° viewport is selected by choosing a region in the viewport while the 90° viewport is presenting. For example, when a participant selects a region of a 90° viewport, the 30° viewport of the selected region is displayed. This likely works a zooming operation. Likewise, the 10° viewport is selected by choosing a region from the 30° viewport. The participants are asked to evaluate each viewport provided for each FoV. The five-level scale is applied as follow:

1 – Very annoying
2 – Annoying
3 – Slightly annoying
4 – Perceptible but not annoying
5 – Imperceptible

When the participant completes the evaluation of all viewports, a summary of rating scores is provided and a final overall score for the Projection X is asked for. An exemplified screen for our DCR test is shown in Fig. 7.

Fig. 8 illustrates the proposed PC test framework. The PC test is performed for all possible pairs of projection formats in each group, which is determined by the DCR test. To compare the quality of the viewport between a Projection X1 and a Projection X2 in a same group, the participant is forced to freely choose and evaluate ten pairs of viewports generated from X1 and X2. An exemplified user interface for this procedure is shown in Fig. 9. The evaluation is performed simply to choose the projection showing a better quality between both displayed viewports. If the difference is transparent, participant may skip a vote and move to the next paired comparison of the viewport at a different PoV. In the last stage of the PC test, the results for all participants are collected and averaged.

IV. EXPERIMENTAL RESULTS

A. SUBJECTIVE TEST SET-UP

We have conducted an experiment that was set-up under the framework proposed in Section III. The proposed DCR and PC tests were performed in order. The DCR test was set-up in three different monitors to discover the effects of the monitor types on the quality rating of the projection formats when the query image is compared to the ground truth image.

FIGURE 5. Six viewports of a 360° image: (a) front, (b) left, (c) back, (d) right, (e) up, and (f) down.

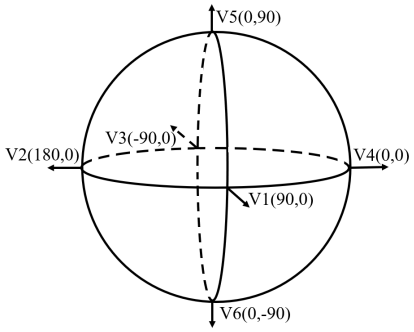


FIGURE 6. Six PoVs of the viewports used in the DCR test.

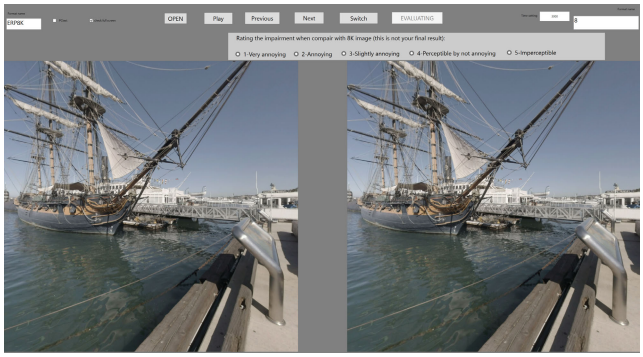


FIGURE 7. DCR test interface.

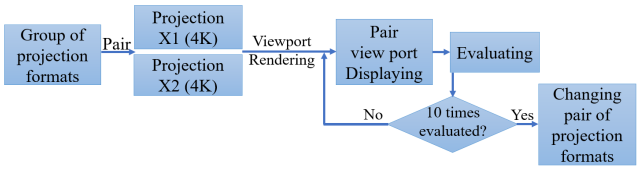


FIGURE 8. The proposed PC test framework.

Two groups of 15 participants have participated in our subjective test. The first group performed the DCR test and the second group performed the PC test. All the participants had normal vision, were non-expert in the field of 360° image processing, had experienced 360° image services, and their ages ranged from 20s to 40s. The participants were required to join a training class and practiced the test programs with sample images before starting their real test. We ensured

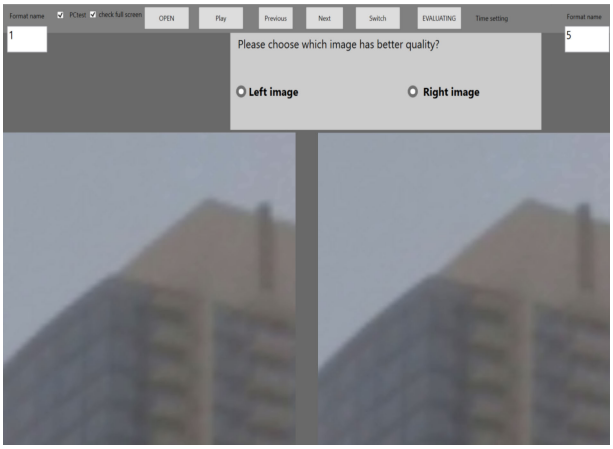


FIGURE 9. PC test interface.

that all participants understood the purpose of the test and how to evaluate the image quality. When the participants agreed to join our experiment, they signed a commitment to perform the test with integrity and to grant the rights to use their results. Both experiments were performed under controlled ambient luminance levels. There were no strongly colored objects around the participant's FoV. The surroundings of the display monitor were set as a gray color to reduce the interaction of the environment on image quality. The selection of our participants follows the ITU-T recommendations [22], [26]. It is noted that the same rules of set-up and participant selection were also successful in several previous experiments [28], [29].

We selected five sample images as shown in Fig. 10. Their 8K ERP versions used as the source images for our experiment. We included outdoor, indoor, and nighttime scenes which were taken from different 360° cameras. These source images were bitmap images for which any lossy encoding was not applied.

The DCR test was conducted on three different displays: 15in, 27in, and 43in, denoted as M1, M2, and M3, respectively. The 15 participants who participated in the DCR test were divided into three groups, with five members in each display group. Each participant in a group performed the experiment on one type of monitor. The 15in and 27in monitors are at 1920 × 1080 resolution, while the 43in monitor

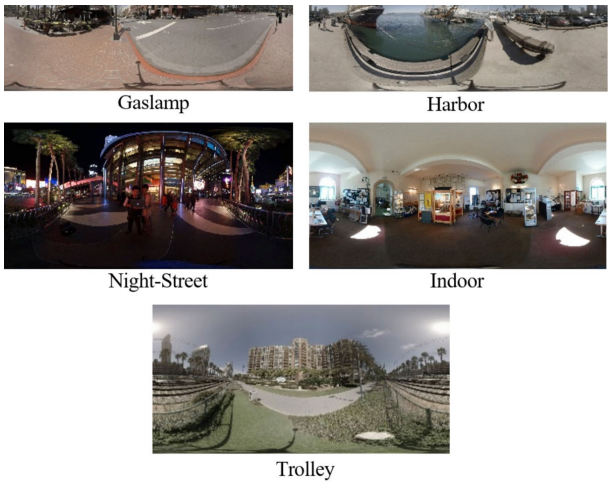


FIGURE 10. Sample images.

is at 3840 × 2160 resolution. All three monitors had a very good color reproduction. We defined the distance from the participant to the monitor at 3 × H, where H is the height of the monitor, as recommended in the ITU-T recommendation P.910 [22] and ITU-R recommendation BT.500-12 [30]. Thus, the viewing distance was 30–40 cm for 15in, 90–100 cm for 27in, and 130–150 cm for 43in. The participants were free to adjust their position within the range of distance for their convenience. The DCR and PC test took about 30 minutes per image, respectively. To avoid participant fatigue, they were allowed a ten-minute break after finishing one image. The PC test was performed after analyzing the results obtained from the DCR test. Based on the DCR test results, the ten 360° projection formats were classified into different groups. All possible combinations between pairs of formats from the ten 360° projection formats in each group were considered, as each pair needed to be evaluated by the PC test. To reduce the time consumed to evaluate and increase the accuracy, the PC test was only performed on the M3 monitor.

B. RESULTS OF DCR TEST

In the DCR test, we collected 5 scenes ×15 participants ×10 projection formats × (3 FoV modes × 6 viewports + overall results) = 14,250 values in total from participants' evaluations. The overall mean opinion score (MOS) for each projection format is calculated and represented in Fig. 11. Fig. 12 shows the MOS of each projection format in the three different monitors. It is apparent from both charts that the TSP format received the lowest quality scores and a group (G1 group) of ACP, EAC, ECP, RSP, and HEC formats always received higher quality scores than the other group (G2 group) of AEP, ERP, CMP, and OHP formats for all three monitors.

Fig. 13 illustrates the DCR test results on three FoV modes. The results show that the quality difference between projection formats is hardly noticeable at FoV = 90°; however,

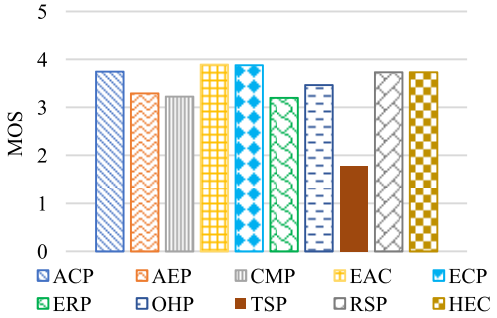


FIGURE 11. Overall MOS results of DCR test.

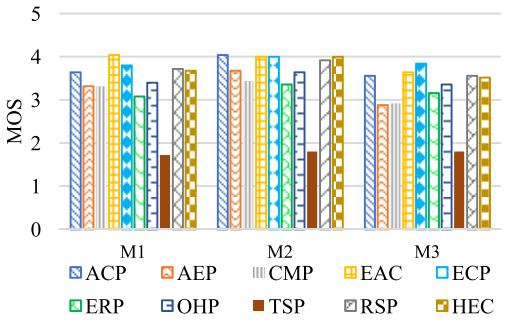


FIGURE 12. Overall MOS of DCR test in different monitor.

the TSP format shows the lowest quality. With FoV = 30°, there is a slightly noticeable difference, with G2 formats having lower quality than G1 formats. At FoV = 10°, the difference between the ten formats is more clearly noticeable. Interestingly, the results for FoV = 10° show that the ranking of the ten formats shows the same trend as FoV = 30°.

The overall MOS score is represented in Fig. 14, and the result coincides with that of FoV = 10°. Owing to these results, we decided to perform our PC test by setting FoV = 10° exclusively.

C. RESULT OF PC TES

The purpose of our PC test is to rank the quality differences of the ten projection formats more precisely. The pair of viewports were displayed randomly. To be fair, all possible pair combinations of projection formats in each group were judged. In the proposed PC test framework, as the participant selects the higher quality viewport between the two shown on the monitor, one point is accumulated in the total score of the selected format. Each participant's results were converted to percentages to ensure that all participants' results have the same weight in the final results. When participants evaluated a pair of formats, there would be m times in which format A was voted and n times in which format B was voted. We calculate the score of projection format A versus projection format B as P calculated by (1).

$$P = \frac{m}{m+n} \tag{1}$$

In this test, each pair of projection formats was evaluated by 15 participants with 5 images. We obtained a total

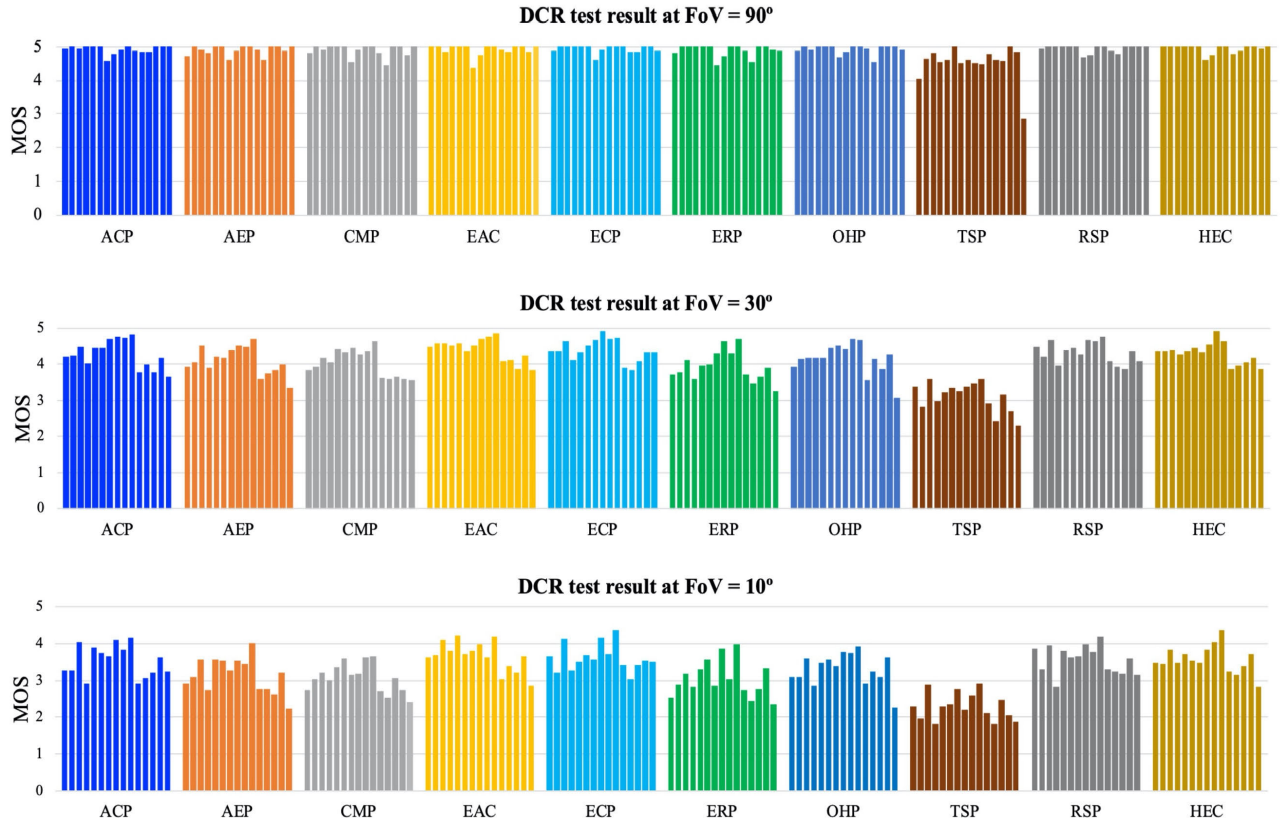


FIGURE 13. MOS of DCR at different FoV (each bar corresponds to one participant).

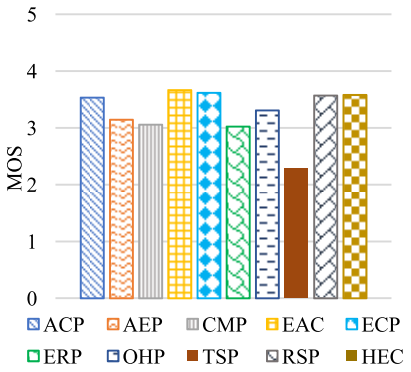


FIGURE 14. Average score of DCR for FoV = 10°.

of 15 participants × 5 images results for each pair of formats. The averaged P values of the formats for the PC test is illustrated in Fig. 15.

The PC test results shows that the RSP format had lower quality than the other formats in G1. The HEC and ECP formats showed a better quality than the ACP and EAC formats, and the quality difference between HEC and ECP was negligible. We can also observe that the ERP format was evaluated as the lowest quality format when compared with the other formats in the G2 group. Further discussions of the results are presented in Section V.

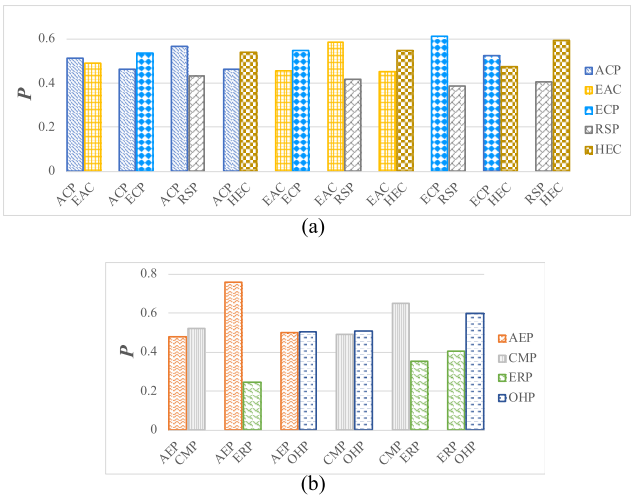


FIGURE 15. PC test results. (a) a result of the group G1 and (b) a result of the group G2.

V. DISCUSSIONS

In this section, we discuss the effect of different monitors on our DCR scoring results in subsection A. We validate our PC test results in section V-B. The one-way analysis of variance (ANOVA) test and T-test are adopted in section V-A and -B, respectively, because we need to validate

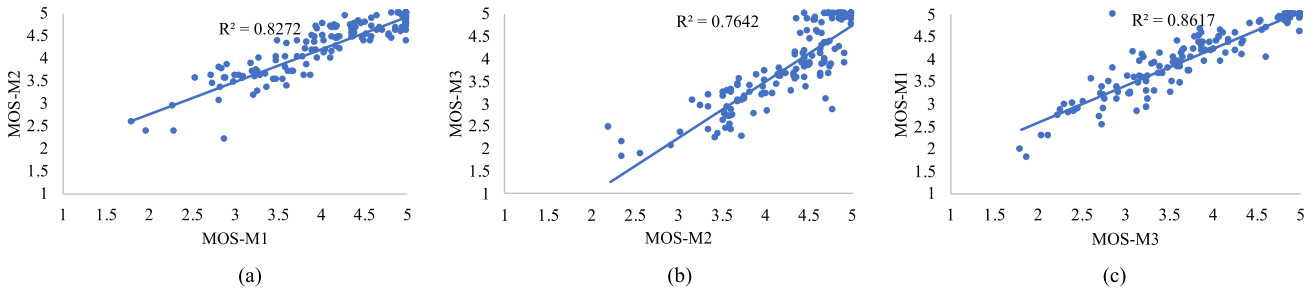


FIGURE 16. Scatter plots of MOS with correlation coefficient of ten projection formats in three monitors. (a) M1 versus M2, (b) M2 versus M3, and (c) M3 versus M1.

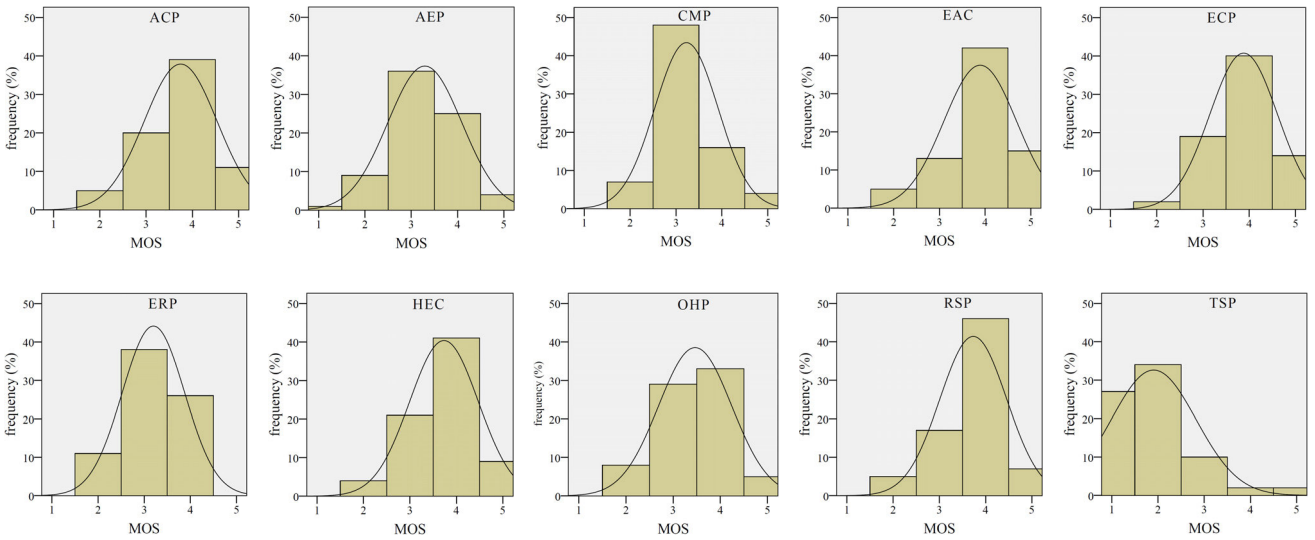


FIGURE 17. Distribution of the MOS for ten projection formats.

the means of three groups and two groups for corresponding cases, respectively [34], [35].

A. COMPARISON OF DCR RESULTS FOR DIFFERENT MONITOR

The 360° image can be used in various devices. We checked if the size of monitors would lead to different results. First, we observed the deviation of DCR scores based on the correlation coefficients between monitors. Fig. 16 is the scatter plots which show the distribution of the MOS of ten projection formats. There are strong linear relationships between monitors. The correlation coefficient (denoted by R^2 in Fig. 16) was between 0.76 and 0.86, which means that the dependency of our results on different monitors is almost negligible [26].

Second, we performed the ANOVA test for further exploring the effect of different monitors on our DCR scores. Fig. 17 presents the distribution of the DCR test result. It is shown that the distribution resembles the normal distribution; hence, usage of the ANOVA test can be justified [31]. We set the null hypothesis H_0 as the assumption that there is no significant difference between participant scores for each projection format in different monitors. Table 3 contains the ANOVA summarization result for each projection format. The "Sig." value denotes the significant difference value calculated for three monitors for each projection format. We determine that H_0 is true if the "Sig." value is greater than the significance level 0.05, which is recommended in [34] and used widely in statistics. All the projection formats, except AEP greater than 0.05, thus we determine that H_0 is true.

TABLE 3. ANOVA summarization for difference between three monitors.

Projection Formats	Sig.	Projection Formats	Sig.
ACP	0.069	ERP	0.327
AEP	***0.001***	OHP	0.392
CMP	0.***018***	TSP	0.513
EAC	0.149	RSP	0.213
ECP	0.601	HEC	0.064

In the case of AEP and CMP, we further performed a post hoc test to observe the more detailed differences [32], [34]. Table 4 contains the post hoc test result for the significant difference of AEP and CMP for three monitor pairs of M1-M2, M1-M3, and M2-M3. It is shown that only the M2-M3 pair

TABLE 4. Multiple comparisons of AEP and CMP at M1, M2, M3.

			Sig.
AEP	M1	M2	0.206
		M3	0.097
	M2	***M3***	***0.001***
CMP	M1	M2	0.797
		M3	0.089
	M2	***M3***	***0.019***

TABLE 5. Shapiro-wilk normality test of PC test score for G1 group.

X1 \ X2	ACP	EAC	ECP	RSP	HEC
ACP		0.9254	0.1477	***0.0348***	0.4691
EAC	0.9254		0.4543	0.2298	0.6704
ECP	0.1477	0.4543		***0.0279***	0.0625
RSP	***0.0348***	0.2298	***0.0279***		0.251
HEC	0.4772	0.6704	0.0625	0.2518	

TABLE 6. Shapiro-wilk normality test of PC test score for G2 group.

X1 \ X2	AEP	CMP	ERP	OHP
AEP		0.1632	***1.65E-05***	0.4731
CMP	0.2308		0.4383	0.1507
ERP	***1.65E-05***	0.1507		0.2136
OHP	0.4731	0.4383	0.2136	

shows the significant difference. Based on these results and noting that AEP and CMP were classified into the same group for our PC test, we may conclude that the impact of different monitors on the DCR quality evaluation of the projection formats is almost negligible for our overall subjective test.

B. ANALYSIS OF THE PC TEST RESULTS

The pc test was separately performed for G1 and G2 groups of projection formats resulting from the DCR test. We first checked the distribution of the scores of the PC test by using the Shapiro-Wilk normality test [31], [33]. Tables 5 and 6 summarize the results for each pair of projections (denoted by X1 and X2) involved in each pc tests of G1 and G2 groups, respectively. It is noted that the data are generally accepted as normally distributed in statistics when the Shapiro-Wilk result is larger than 0.05 [31]. It is shown that most pairs, except RSP-ACP, RSP-ECP, and ERP-AEP pairs, pass the Shapiro-Wilk normality test.

We performed T-test for each pair of projection formats involved in each PC test, and the results are presented in Table 7. The ''Sig'' column presents the significant difference level of the tested pair, and it is assumed that there is no significant difference if the ''Sig'' value is less than a typically used threshold value of 0.05 in statistics [35].

TABLE 7. T-test results for each pair of formats in the PC test.

Group	Pair	Pair of the projection formats		Mean difference (X1 - X2)	Sig.
		(X1)	(X2)		
G1	Pair 1	ACP	EAC	0.023	0.609
	Pair 2	ACP	ECP	-0.070	0.073
	Pair 3	ACP	RSP	0.135	***0.002***
	Pair 4	ACP	HEC	-0.075	0.089
	Pair 5	EAC	ECP	-0.091	***0.034***
	Pair 6	EAC	RSP	0.168	***0.000***
	Pair 7	EAC	HEC	-0.095	***0.019***
	Pair 8	ECP	RSP	0.224	***0.000***
	Pair 9	ECP	HEC	0.050	0.180
	Pair 10	RSP	HEC	-0.188	***0.000***
G2	Pair 11	AEP	CMP	-0.042	0.332
	Pair 12	AEP	ERP	0.511	***0.000***
	Pair 13	AEP	OHP	-0.003	0.950
	Pair 14	CMP	OHP	-0.018	0.703
	Pair 15	CMP	ERP	0.297	***0.000***
	Pair 16	ERP	OHP	-0.195	***0.000***

Based on Table 7, we may conclude the following for the projection formats belonging to the G1 group.

- There are six cases (pair 3, pair 5 to pair 8, and pair 10) showing that the significant difference is less than 0.05. Among these cases, four are the pairs with RSP and two other cases are the pairs OF EAC-ECP and EAC-HEC.
- Even though the pairs with RSP show the failure of Shapiro-Wilk normality test and show no significant difference result in the T-test, RSP is evaluated as having the lowest quality in group G1 because the mean scores of RSP are much smaller than the paired formats.
- Except the pair including RSP, the mean difference of each pair is small. However, if we focus on the pairs whose significant value is above the threshold value, we may decide that ACP is better than EAC (pair 1), both ECP (pair 2) and HEC (pair 4) are better than ACP, and ECP is better than HEC (pair 9).
- Finally, we may rank the five formats in G1 from highest to lowest quality as ECP, HEC, ACP, EAC, and RSP even though the quality difference between ECP, HEC, ACP, and EAC is small.

The analysis for G2 in the PC test is given as follow:

- There are three cases (pairs 12, 15, and 6) showing that the significant difference is less than 0.05. interestingly, all those cases are the pairs with ERP, and the mean score of the ERP is lower than the paired format in all cases.
- For other pairs in G2, there is no other case where the significant difference is less than 0.05. therefore, we may rank the quality as OHP, CMP, AEP, and ERP in descending order.

As a result, our ranking of the ten projection formats from the lowest to the highest quality is summarized as follows:

1. The lowest quality format was TSP.
2. The ERP format has lower quality than the other formats, with the exception of TSP.
3. Three formats, AEP, CMP, and OHP, follow in order of quality ranking; however, the difference in quality is minimal.
4. The RSP lies in the higher quality ranked G1 group; however, it has the lowest quality in this group.
5. The ACP and EAC formats showed lower quality than HEC and ECP formats in the G1 group.
6. The ECP and HEC formats were evaluated as having the highest quality of the ten projection formats considered in this study.

VI. CONCLUSION

In this paper, we presented an overview of the 360° image and the well-known subjective test methods for image quality evaluation and proposed a framework for the 360° image's subjective evaluation through its viewports. The combination of the DCR and PC test methods in the proposed framework will help to classify the quality of different projection formats. We conducted the experiments following the proposed methodology. The projection formats were evaluated by adjusting the FoV. We obtained the DCR test result at three different monitors to evaluate the effect of different monitors on the evaluation of the ten projection formats. The results reveal that different monitors negligibly affect the quality ranking of the ten projection formats and show the ECP and HEC are of outstanding formats among ten projection formats as far as visual quality is concerned.

The results can be used as a reference for future work related to selecting a high quality projection format for the 360° image, particularly for selecting a projection format for the 360° image standard. Because of the expense and time consuming nature of the subjective test, we performed the test with a limited number of image sources and participants. However, the evaluation results of the participants were very similar; therefore, we are confident in our conclusions.

It would be also better to mention that the results of subjective image quality evaluation are often used for developing objective image quality metrics. Most existing objective quality metrics are trained by nonlinear regression or rank learning [36]–[38]. They use either one of absolute or relative quality scores obtained by subjective tests. However, our evaluation methodology jointly considers the absolute and relative quality using DCR and PC, respectively, and there are rare objective quality metrics combining the absolute and relative scores of subjective tests together. It is suggested as a further research topic to develop a no-reference objective image quality metric trained by our combined subjective test results.

REFERENCES

[1] P. Zimmermann, "Virtual reality aided design. A survey of the use of VR in automotive industry," in *Product Engineering: Tools and Methods Based on Virtual Reality*, D. Talaba and A. Amditis, Eds. Dordrecht, The Netherlands: Springer, 2008, pp. 277–296.

[2] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shiban, and A. Mühlberger, "The impact of perception and presence on emotional reactions: A review of research in virtual reality," *Frontiers Psychol.*, vol. 6, p. 26, Jan. 2015.

[3] A. Bhat, G. Bhagwat, and J. Chavan, "A survey on virtual reality platform and its applications," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 4, no. 10, pp. 3775–3778, 2015.

[4] J. Ardouin, A. Lécuyer, M. Marchal, and E. Marchand, "Navigating in virtual environments with 360 omnidirectional rendering," in *Proc. IEEE Symp. 3D User Int. (3DUI)*, Mar. 2013, pp. 95–98.

[5] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan, "Assessing visual quality of omnidirectional videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3516–3530, Dec. 2019.

[6] X. Xiu, Y. He, Y. Ye, and B. Vishwanath, "An evaluation framework for 360-degree video compression," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

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

[8] H. T. Tran, C. T. Pham, N. P. Ngoc, A. T. Pham, and T. C. Thang, "A study on quality metrics for 360 video communications," *IEICE Trans. Inf. Syst.*, vol. 101, no. 1, pp. 28–36, 2018.

[9] *Working Draft 3 of Versatile Video Coding*, Standard ISO/IEC JTC1/SC29/WG11 N18027, 2018.

[10] *Draft Call for Proposals on JPEG 360 Metadata*, Standard ISO/IEC JTC 1/SC29/WG1N77012, 2017.

[11] *Information Technology—Coded Representation of Immersive Media—Part 2: Omnidirectional Media Format*, Standard ISO/IEC FDIS 23090-2, 2018

[12] *SVN_360Lib*. Accessed: Jul. 22, 2019. [Online]. Available: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/trunk/

[13] Y. Sullivan, *JVET 360Lib Software Manual*, document ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Joint Video Exploration Team, ITU-T, Geneva, Switzerland, 2016. Accessed: Jul. 22, 2019. [Online]. Available: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/trunk/

[14] Y. Ye, E. Alshina, and J. Boyce, *Algorithm Descriptions of Projection Format Conversion and Video Quality Metrics in 360Lib*, document ITU-T SG 16 WP 3 and ISO/IECJTC1/SC29/WG11, Joint Video Exploration Team, vol. 16, 2017.

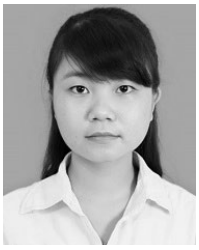
[15] M. Zhou, *AHG8: A Study on the Quality Impact of Line Re-Sampling Rate in EAP*, Standard ISO/IEC JTC1/SC29/WG11, JVET-F0025 and Joint Video Exploration Team, ITU-T SG16 WP3, 2017.

[16] M. Zhou, *AHG8: A Study on Equi-Angular Cubemap Projection (EAC)*, document ITU-T SG 16 WP 3 ISO/IECJTC1/SC29/WG11, Joint Video Exploration Team, 7th Meeting: JVET-G0056, Turin, Italy, 2017.

[17] Y.-H. Lee, J.-L. Lin, S.-K. Chang, and C.-C. Ju, *CE13: Modified Cubemap Projection in JVET-J0019*, document ITU-T SG16 WP3ISO/IEC JTC1/SC29/W11, JVET-K0131, Joint Video Exploration Team, 2018.

[18] G. Van der Auwera, M. Coban, M. Hendry, and M. Karczewicz, *AHG8: Truncated Square Pyramid Projection (TSP) for 360 Video*, document ITU-T SG16 WP3 ISO/IEC JTC1/SC29/W11, JVET-D0071, Joint Video Exploration Team, 2016.

[19] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, "Comparison of four subjective methods for image quality assessment," *Comput. Graph. Forum*, vol. 31, no. 8, pp. 2478–2491, Dec. 2012.
[20] M. H. Pinson, M. Barkowsky, and P. Le Callet, "Selecting scenes for 2D and 3D subjective video quality tests," *J. Image Video Process.*, vol. 2013, no. 1, Dec. 2013.
[21] M. H. Pinson, L. Janowski, and Z. Papir, "Video quality assessment: Subjective testing of entertainment scenes," *IEEE Signal Process. Mag.*, vol. 32, no. 1, pp. 101–114, Jan. 2015.
[22] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document ITU-T P.910, ITU-T, 2008.
[23] T. Kawano, K. Yamagishi, and T. Hayashi, "Performance comparison of subjective assessment methods for stereoscopic 3D video quality," *IEICE Trans. Commun.*, vol. E97.B, no. 4, pp. 738–745, 2014.
[24] T. T. H. Uyen, S. Choi, and O.-J. Kwon, "Subjective evaluation of the 360-degree projection formats using absolute category rating," in *Proc. ICITT*, Da Nang, Vietnam, 2019, pp. 10–15.
[25] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Sep. 2015, pp. 31–36.
[26] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*, document ITU-T P.913, ITU-T, 2016.
[27] *Viewing Conditions-Graphic Technology and Photography*, document ISO 3664:2000, 2000.
[28] S. Youn, S. Baek, T. Jeong, and C. Lee, "Perceptual video quality comparison of various 3D video formats and displays," *Displays*, vol. 52, pp. 21–29, Apr. 2018.
[29] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, "Evaluation of HDR tone mapping methods using essential perceptual attributes," *Comput. Graph.*, vol. 32, no. 3, pp. 330–349, Jun. 2008.
[30] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document BT.500-12, ITU-R, 2012.
[31] A. Mayers, "Normal distribution," in *Introduction to Statistics and SPSS in Psychology*. Harlow, U.K.: Pearson, 2013, pp. 37–62.
[32] A. M. Brown, "A new software for carrying out one-way ANOVA post hoc tests," *Comput. Methods Programs Biomed.*, vol. 79, no. 1, pp. 89–95, Jul. 2005.
[33] N. M. Razali, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *J. Stat. Model. Anal.*, vol. 2, no. 1, pp. 21–33, 2011.
[34] A. Mayers, "Independent one-way ANOVA," in *Introduction to Statistics and SPSS in Psychology*. Harlow, U.K.: Pearson, 2013, pp. 170–190.
[35] A. Mayers, "Independent t-test," in *Introduction to Statistics and SPSS in Psychology*. Harlow, U.K.: Pearson, 2013, pp. 1137–1153.
[36] J. Kim, "Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 130–141, Nov. 2017.
[37] X. Liu, J. V. D. Weijer, and A. D. Bagdanov, "RankIQA: Learning from rankings for no-reference image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1040–1049.
[38] Q. Wu, H. Li, Z. Wang, F. Meng, B. Luo, W. Li, and K. N. Ngan, "Blind image quality assessment based on rank-order regularized regression," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2490–2504, Nov. 2017.



TRAN THI HAI UYEN received the bachelor's degree in electronics engineering from the University of Transport and Communication, Vietnam, in 2014. She is currently pursuing the M.S. degree in electronics engineering with Sejong University, Seoul, South Korea. She was an Engineer with the Samsung's Vietnam Mobile Research and Development Center, from 2014 to 2017. Her research interests include image and video processing, and 360° image and video coding, processing, and analyzing.

OH-JIN KWON received the M.S. degree from the University of Southern California, Los Angeles, CA, USA, in 1991, and the Ph.D. degree from the University of Maryland, College Park, MA, USA, in 1994, both in electrical engineering. He was a Researcher at the Agency for Defense Development of Korea, from 1984 to 1989, and the Head of the Media Lab, Samsung SDS Company, Ltd., Seoul, South Korea, from 1995 to 1999. Since 1999, he has been a Faculty Member at Sejong University, Seoul, where he is currently a Professor. His research interests include image and video fusion, coding, watermarking, analyzing, and processing.

SEUNGCHEOL CHOI received the B.S. and M.S. degrees in computer science from Sejong University, Seoul, South Korea, in 1998 and 2001, respectively, and the Ph.D. degree in electronics engineering from Sejong University, in 2017. He was a Researcher at Galaxia Communications, from 2001 to 2013. He has been a Postdoctoral Researcher at Sejong University, since 2017. His research interests include image and video coding, high-dynamic-range imaging, image processing, image fusion, and JPEG.

IKRAM HUSSAIN received the bachelor's degree in computer system engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2016. He is currently pursuing the M.S. degree in electronics engineering with Sejong University, Seoul, South Korea. His research interests include image and video processing, as well as 360° image and video coding, processing, and analyzing.