Subjective Panoramic Video Quality Assessment Database for Coding Applications

Yingxue Zhang, Yingbin Wang, Feiyang Liu, Zizheng Liu, Yiming Li, Daiqin Yang, *Member, IEEE*, Zhenzhong Chen, *Senior Member, IEEE*

***Abstract*—With the development of virtual reality, higher quality panoramic videos are in great demand to guarantee the immersive viewing experience. Therefore, quality assessment attaches much importance to correlated technologies. Considering the geometric transformation in projection and the limited resolution of head-mounted device (HMD), a modified display protocol of the high resolution sequences for the subjective rating test is proposed, in which an optimal display resolution is determined based on the geometry constraints between screen and human eyes. By sampling the videos to the optimal resolution before coding, the proposed method significantly alleviates the interference of HMD sampling while displaying, thus ensuring the reliability of subjective quality opinion in terms of video coding. Using the proposed display protocol, a subjective quality database for panoramic videos is established for video coding applications. The proposed database contains 50 distorted sequences obtained from ten raw panoramic video sequences. Distortions are introduced with the High Efficiency Video Coding compression. Each sequence is evaluated by 30 subjects on video quality, following the absolute category rating with hidden reference method. The rating scores and differential mean opinion scores (DMOSs) are recorded and included in the database. With the proposed database, several state-of-the-art objective quality assessment methods are further evaluated with correlation analysis. The database, including the video sequences, subjective rating scores and DMOS, can be used to facilitate future researches on coding applications.**

***Index Terms*—Virtual reality, panoramic video, subjective video quality database, video quality assessment, video coding.**

I. Introduction

VIRTUAL Reality (VR) has attracted much attention and massive effort has been put into related researches recently [1]–[3]. VR presents simulated and immersive environment for users with a Head-Mounted Device (HMD) displaying panoramic, stereoscopic videos [4]. Panoramic video, one of the most primary media contents of VR, can provide 360° free viewing for the viewers with a HMD and differs greatly from the conventional, 2D videos displayed with a flat screen [5].

The establishment of immersive scene calls for higher quality and faster transmission of panoramic videos. However, the compression before videos reaching the viewers mostly brings impairments to the videos. The degradation may greatly influence the Quality of Experience (QoE). Therefore, quality assessment methods for panoramic videos are essential in the future development of VR and related technologies.

Both objective and subjective methods are used for quality assessment. Objective methods assess the video quality with mathematical models, which can be conducted automatically without too much human involvement. With this advantage, objective quality assessment has been well investigated [6]–[8]. There are also some objective quality assessment methods being already applied to evaluate the quality of impaired panoramic videos [9]–[11]. For example, a Sphere based Peak Signal-to-Noise Ratio (S-PSNR) method based on the PSNR being widely used in quality evaluation is proposed in [10]. The S-PSNR estimates the average quality over all viewing directions by computing PSNR with the uniform sample points on the sphere, rather than on the mapped plane. Additionally with the statistical results on the head motion data, a latitude-weighted S-PSNR is further proposed with the relative viewing frequencies on corresponding latitudes of the sphere to better approximate the quality of viewport.

Despite the convenience of the objective models, subjective quality assessment is the most direct and reliable method to know how the observers feel about the video quality. Therefore, the development of subjective quality assessment is of much necessity for evaluating the video quality and benchmarking the performance of objective models. Organizations such as International Telecommunication Union (ITU) [12], Video Quality Experts Group (VQEG) [13] have recommended varieties of subjective assessment test plans on images and videos to evaluate different aspects of multimedia processing systems. Protocols differ significantly given different test aims and targets to be evaluated.

Under this circumstance, many works on subjective quality assessment are conducted for varied applications, such as the LIVE video quality database [14] for 2D video,

Manuscript received August 21, 2017; revised November 24, 2017; accepted January 16, 2018. Date of publication April 26, 2018; date of current version June 5, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61471273 and Grant 61771348, and in part by the Wuhan Morning Light Plan of Youth Science and Technology under Grant 2017050304010302. *(Corresponding author: Zhenzhong Chen.)*

Y. Zhang, Y. Wang, F. Liu, Z. Liu, and D. Yang are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China.

Y. Li is with the School of Computer Science, Wuhan University, Wuhan 430072, China.

Z. Chen is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China, and also with the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: zzchen@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2018.2811627

NAMA3DS1-COSPAD1 for stereoscopic sequences [15], Lutz Goldmann *et al.*'s work for QoE of 3D videos [16], LIVE Mobile Video Quality Database for mobile video over wireless network [17], and Jooyoung Lee *et al.*'s work for mixed-resolution stereoscopic video broadcasting service [18].

However, few subjective databases for panoramic video are publicly available currently. With unique characteristics and being a new media content, panoramic videos call for specified subjective quality assessment databases to help figure out the quality perceived by the viewers, which motivates the improvements on quality of experience and coding systems.

Moreover, during subjective quality assessment, an ideal playback system that do not bring extra artifacts to the test videos is essential [19]. In natural video quality assessment, the test sequences with coding artifacts are usually displayed in a per-pixel manner in order to avoid any other quality change. Therefore, we also try to guarantee a per-pixel presentation of the panoramic video during the test. However, as we know, the Field of View (FoV) and resolution of the HMD screen is limited, while the panoramic videos are commonly at a high resolution. Taking HTC VIVE as an example, corresponding to the FoV of 110°, the horizontal display resolution of 8K test sequences will be approximately 2503, while the horizontal resolution of the screen is 1080. This mismatch inevitably leads to sampling of the videos, hiding or amplifying the coding artifacts to be evaluated in the subjective quality assessment. For example, the blocking artifact from compression may not be perceived if the video is down-sampled and the artifact will, conversely, be more noticeable if up-sampling is conducted.

The interference, therefore, misleads the perception of video quality. Since the subjective quality assessment of panoramic videos is a newly emerging topic, there are few existing researches that can be refereed to on how to ensure the per-pixel display of the panoramic video. Currently, existing studies conduct subjective experiments with HMDs directly, ignoring the influence brought by complicated optical system and projection while compressing and displaying, which do not meet the requirement of subjective quality assessment and calls for further discussion.

Concerning all the aforementioned problems, a subjective quality database for panoramic videos on coding conditions is established in this paper. During the subjective test, the idea of re-sampling the video to an optimal resolution before coding is proposed and is believed to be able to alleviate the interference of the HMD display for subjective assessment. The subjective quality assessment test is conducted based on 60 high-resolution panoramic videos with coding degradations. The Absolute Category Rating with Hidden Reference (ACR-HR) method [20] is utilized. 30 subjects participate in the rating test towards video quality. Furthermore, the performance of existing objective quality assessment methods is also evaluated using the proposed database.

Rest of the paper is arranged as follows. Section II introduces some related work on quality assessment. Section III discusses the sampling problem of panoramic video display and optimal display protocol for subjective quality assessment. Section IV describes the process of establishing the proposed subjective quality database. Section V discusses the processing of the subjective quality rating data and makes a summary of the database. Section VI shows the performance evaluation of some existing objective quality assessment models based on the database. Section VII makes a conclusion for the paper.

II. Related Work

Considering the significance of subjective quality assessment in image and video applications, there have been various approaches obtaining subjective quality opinions and establishing databases for different purposes. In this section, the protocols for subjective rating tests, existing subjective quality databases and problems on sequence presentation during subjective test are reviewed.

A. Protocols for Subjective Quality Assessment

The protocols that focus on the subjective quality assessment of videos have been well studied. The assessment can be generally classified into two categories, i.e., the methods using Single Stimulus (SS) and Double Stimulus (DS). Different methods are chosen according to specific aim of the test. In this section some representative methods are described and more possibilities can also be found in related recommendations and reports like [20]–[22], etc.

When the test aims at figuring out the absolute quality of the videos, Single Stimulus (SS) method can be used by showing each sequence independently to the subject. One of the representative method is Single Stimulus Continuous Quality Evaluation (SSCQE) [21] that is proposed to handle the time-varying quality of coded videos by allowing subjects to rate dynamically on the quality [23]. Compared with SSCQE, Absolute Category Rating (ACR) method [20] is easier to implement and suitable for qualification tests in that its presentation protocol is similar to the common use of the system.

Owning the advantage of ACR, the ACR-HR method, being efficient and easy, shows additional superiority. Though the sequences are displayed one at a time, the bias of scene contents and reference quality can be removed by including all the references in the rating procedure without any special identification. By rating independently on both test sequences and references, the quality with respect to the reference can be figured out by calculating the difference between rating scores, avoiding the impact of video content.

Double Stimulus (DS) method shows two stimuli side by side or continuously in pairs for the subjects at the same time so that the subjects can directly judge difference on the quality or fidelity between the two videos. To evaluate the fidelity of the system, Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) [21] is proposed on the basis of the SSCQE method. With the reference shown beside the test sequence, the subjects are asked to rate the fidelity of the test sequences with respect to the references. Degradation Category Rating (DCR) method [20], on the other hand, is used for impairment evaluation, i.e., whether and to which level the subjects can perceive the impairments.

B. Subjective Quality Assessment

Booming technologies have brought us various types of media contents. Based on the subjective studies mentioned above, many subjective quality databases are established for different kinds of contents and applications.

LIVE video quality database is one of the most widely used 2D video subjective quality database proposed in [14]. A subjective study was conducted to evaluate the effects of representative video compression and communication technologies. Each video was assessed by 38 human subjects. SSCQE procedure was used and Differential Mean Opinion Scores (DMOS) were recorded. As media technologies and user demands develop, the resolution of video contents increases significantly, and the coding, transmission and broadcasting of stereoscopic video also prospers rapidly, the subject perception of the videos at higher resolutions or of three dimensions are well investigated [15], [24]. In addition to the video quality, massive effort has been put into the research on Quality of Experience (QoE), which aims at investigating and improving QoE of 2D or 3D videos [16], [25]. Concerning image quality assessment, there are also many targeted databases for various applications and image types like image retargeting quality [26], full-reference image quality [27], Screen Content Images (SCIs) quality [28], etc.

On the subjective quality assessment methods for panoramic videos, there have been heated discussions, mainly focusing on the viewport based assessment methods, which divides the panoramic video into several parts based on the viewports and presents via flat screen stepwise, as the 2D video assessment does. For example, In [29], four viewport selecting strategies were suggested to be combined together for the subjective comparison of video quality under different projection formats. A subjective test pilot study based on the protocol was conducted in [30] to see the impact of the discontinuous edges from projections. The test was conducted with dynamic viewports rendered from decoded bit stream and presented on the LCD TV monitor. Two pre-rotation methods handling the discontinuous edges were tested. Moreover, due to the different FoV of the viewport of the 360° video from that of 2D Standard-Dynamic Range (SDR) contents, the optimal viewing distance for the viewport based subjective assessment was also proposed in [31].

Though being easy to conduct, the viewport based method ignores some essential features of panoramic videos, which makes it unreliable to reflect the subjective perception. Different from these methods, we propose to allow the subjects to view the video with HMD, ensuring the compact and real view experience.

In addition, the subjective quality data is increasingly important for the validation of related technologies. In [9], a subjective quality evaluation of panoramic videos was mentioned to verify the proposed objective metric, which compared Equirectangular projection (ERP) and Craster projection. In [32], a tiling method for interactive panoramic systems was proposed to reduce bandwidth requirement and improve quality of experience. For validation, subjective evaluation was conducted with ACR scale specified for multimedia applications in [22]. In [33], a subjective test plan for panoramic videos was introduced. Instead of display problem, the influence of different viewing patterns on the observers' psychophysical viewing experience was specially discussed to promote the reliability of the subjective rating scores, and the test plan was described in detail in [34]. As one of the most basic problems that influence the QoE, the viewing discomfort has also been widely studied, e.g., [35]–[38].

Concerning the subjective quality assessment database focusing on the visual quality, however, little work has been done to specifically study the quality perception of panoramic videos associated with coding impairments or establish corresponding databases. With the increasing demand of benchmarking and improving related technologies, e.g., objective quality assessment models and coding systems, a large-scale subjective quality assessment database for panoramic videos is of great necessity.

C. Display Protocol of Panoramic Sequences With VR HMDs

In subjective quality assessment, the proper presentation of test sequences is essential to the reliability of evaluation. The subjective rating will be disturbed if unexpected change on quality is introduced to the sequences by the presentation method as mentioned before. Therefore, the videos are supposed to be presented in a per-pixel manner without any sampling to guarantee least quality change other than coding impairments to be evaluated. Unlike 2D video assessment, in which the problem can be handled by either adding black blocks around the border of the video if the resolution of the video is lower than the screen or showing parts of the video separately if the resolution is much higher, the assessment of panoramic video has to be presented in a full-screen mode to guarantee the spherical, immersive experience. Therefore, sampling is inevitable for panoramas displayed via HMDs. Thus, how to realize a per-pixel presentation makes one of the ultimate problem for subjective quality assessment of panoramic videos.

State-of-the-art studies like [9], [32], and [33] stated before directly displayed the panoramic videos with HMDs without mentioning the sampling and rendering of the videos so that it is hard to determine the validity of the rating data obtained from the experiments.

III. Rendering and Display of Panoramic Sequences for Subjective Quality Assessment

A. Display of Panoramic Videos With HMDs

Panoramic video is one of the most important media content in virtual reality, which refers to the video that contains continuous contents in all directions. Generally, scenes on different directions are captured separately based on certain geometry constraints and stitched together to form a panoramic scenario. To present the 360° scene to the viewers, the video is rendered assuming the human head at the center and displayed using the plane screens in the HMD, with a set of lens recovering the geometry structure of the scene captured.

Considering the limitation on transmission and devices, the panoramic videos have to be compressed before they can be

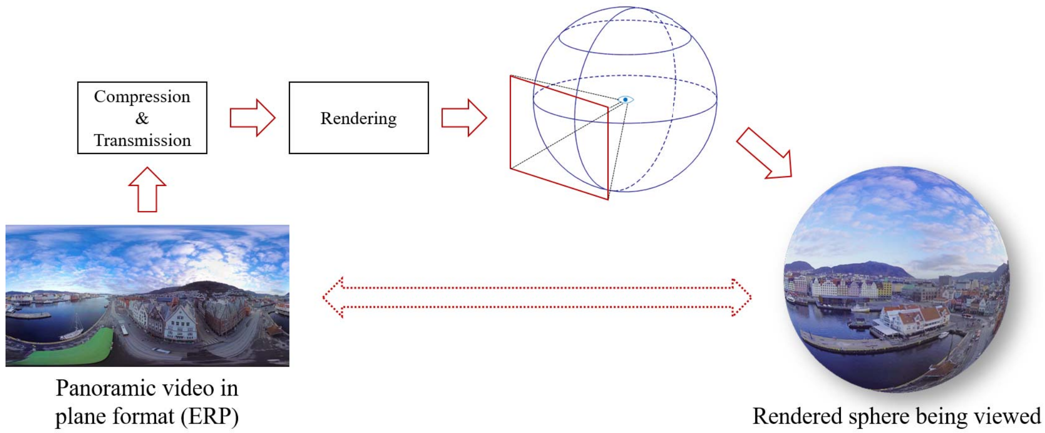


Fig. 1. Panoramic video processing and display.

presented to the viewers, which makes up the main motivation of the proposed subjective quality assessment. As shown in Fig. 1, since existing coding systems cannot be applied to videos in sphere format, the panoramic videos must first be mapped onto a plane in accordance with certain geometric transformation rules, e.g., Equi-rectangular projection (ERP) [39], Cube Map projection (CMP) [39], Icosahedral projection (ISP) [40]. The compressed plane video will again be rendered into a sphere in the aforementioned way while displaying to viewers.

B. Problem on Subjective Quality Assessment of Panoramic Videos

To retain the immersive characteristics and guarantee a precise perception of the quality of panoramic videos, the subjective quality assessment should be conducted with virtual reality HMDs rather than the plane screen monitors. Immersion requires that the virtual content can fill the entire FoV in HMD. Fig. 2 shows the FoV of human eyes in natural and virtual reality viewing conditions. In order to bring immersive experience to the viewers, the FoV α_L' and α_R' of the HMD must keep fixed and consistent with that of the human eyes (shown as α_L and α_R). Otherwise the viewing experience will differ from the natural viewing condition or even bring sickness to the viewers. Unlike the display of 2D video, which can be presented in a per-pixel manner on the screen with fixed size by adding black pixels to the low resolution video or showing only part of the high resolution content, the panoramic video must be presented in its entirety despite of different resolutions. Therefore, the fixed FoV, transformation between plane and sphere, and the multiple resolution of the panoramic videos together lead to the aforementioned sampling problem. The videos of different resolutions have to be up- or down-sampled to a variable extent to fit for the same range of visibility of the HMD. As indicated in Fig. 3, the number of sampling points on a panorama is twice the number on the video half its resolution. Apparently, the sampling of the HMDs may affect the visibility of coding impairments. For example, the blocking artifact may not be perceived by the observers if the video is down-sampled by the HMD, which will mislead the subjective quality perception on the video and results in a higher rating score.

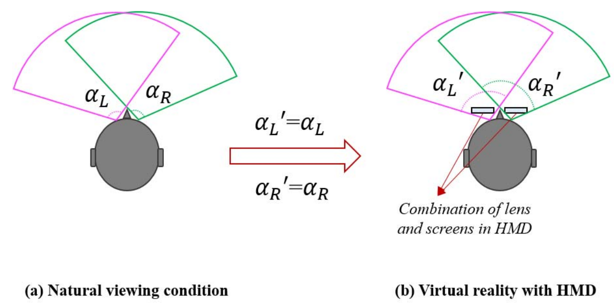


Fig. 2. The relationship of the FoV in natural and virtual reality viewing conditions.

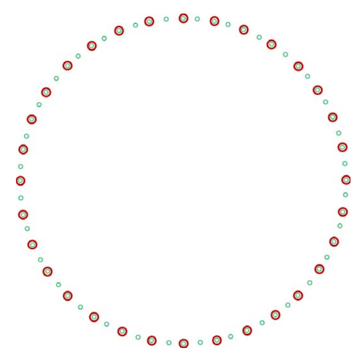


Fig. 3. Example of the density of sampling points on panoramas of different resolutions. Green points (smaller ones) show the sampling points of the video with higher resolution, whereas red points (larger ones) the lower.

C. The Optimal Display Resolution for Subjective Quality Assessment

Based on the above analysis, it can be seen that, in subjective quality assessment test, the test videos cannot be directly compressed and displayed with the HMD in order to avoid unnecessary quality change due to the sampling of different levels. The overlap of sampling problem will interfere the subjects opinion on the video quality in terms of compression. Therefore, despite of the different resolutions of original

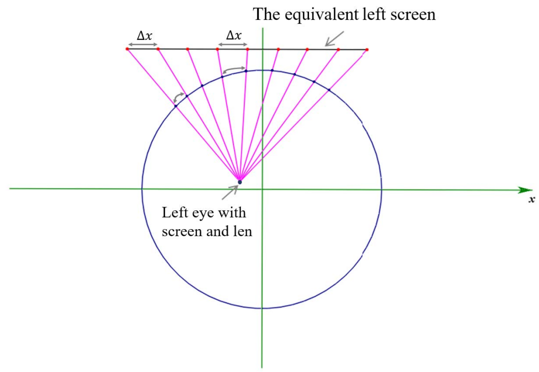


Fig. 4. Perspective projection of panoramic video illustrated with the left screen and eye.

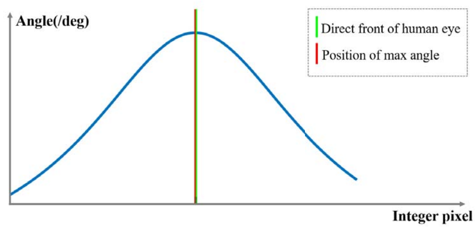


Fig. 5. The angle value distribution between every two adjacent integral pixels on the screen.

panoramic videos, they must firstly be sampled to an uniform and optimal resolution with respect to a certain HMD, which guarantees least quality change while displaying. Then the optimally sampled video can be sent into the coding systems to be evaluated to generate test sequences. Therefore, what we can do to address the sampling problem in subjective rating experiments is to figure out an optimal display resolution for the HMD used in the experiment to guarantee least sampling degradation while presenting to the viewers.

The geometric relationship on the equator of the virtual sphere is shown in Fig. 4, in which red points means the pixels on the screen and Δx shows the constant interval between adjacent pixels. The cluster of lines connecting the left eye and each pixel intersects the equator on a set of sampling points that will finally be projected onto the integral pixel positions of the screen. Provided that the left end of the screen is (x_o, y_o) and the position of left eye is (x_l, y_l), the cluster of lines can be presented as:

$$Y = \frac{y_o - y_l}{m\Delta x + x_o - x_l}(X - x_l) + y_l \tag{1}$$

where the X and Y denote the coordinates of intersection points between the lines and the sphere, which can be determined with:

$$X^2 + Y^2 = r^2 \tag{2}$$

where r represents the radius of the sphere and is empirically set to 12.915 in accordance with the VR HMD used in the tests. The points whose vertical coordinate is greater than zero are determined to be the positions of the sampling points on the equator.

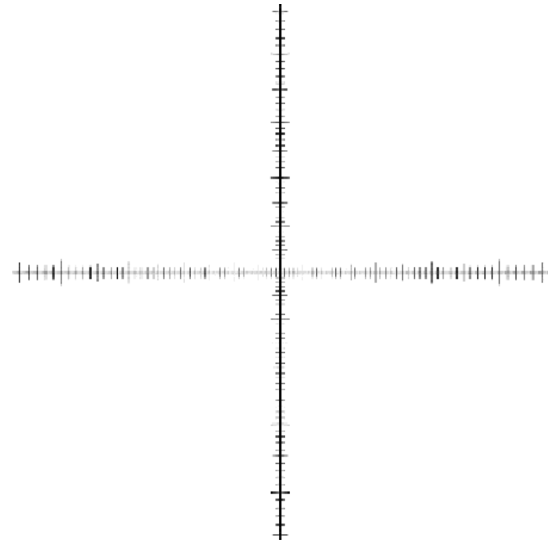


Fig. 6. A standard scale figure displayed by the HMD. The scales being equally spaced on the axis are stretched to be uneven due to projection.

After obtaining the coordinates of the sampling points, the angle between the lines crossing zero point and the n^{th} sampling point is calculated with the radius and the horizontal coordinate X_n:

$$\alpha_n = \sin^{-1}\left(\frac{X_n}{r}\right) \tag{3}$$

Then the angle $\Delta\alpha$ between the adjacent sampling points is figured out:

$$\Delta\alpha = \alpha_n - \alpha_{n-1} \tag{4}$$

Fig. 5 shows the distribution of the $\Delta\alpha$ on the screen, from which we can see that the sampling points on the sphere are not uniformly mapped onto the screen due to perspective projection. The angle between adjacent points decreases as they approximate the edge of the screen, resulting in the stretching effect shown in Fig. 6, which indicates that the sampling will inevitably introduce interference to the viewing experience. Under this circumstance, we aim to guarantee a maximized area on the center of viewport to be presented without sampling, considering the observers' visual tendency towards center. Therefore, the optimal horizontal resolution is defined as:

$$W = \frac{360}{\Delta\alpha_{mid}} \tag{5}$$

where W means the horizontal resolution and the vertical resolution can be calculated with the constraints of specific coding system. $\Delta\alpha_{mid}$ means the angle between center point on the screen and its adjacent one.

Note that the geometry of lens is not introduced into the derivation process. As shown in Fig. 7, a larger FoV can be achieved using a smaller screen with the help of the lens, which significantly reduces the size of HMD. Though the lens additionally brings geometric and color deviations, the rendering system will rectify them and thus presenting consistent scene with that before going into the lens. Therefore, we do

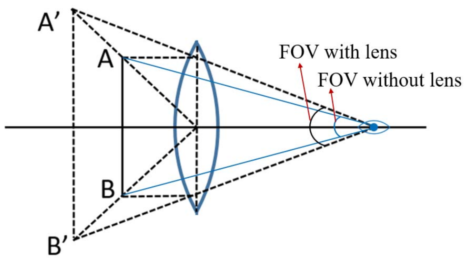


Fig. 7. The optical path through lens in HMD.

not take extra consideration on the geometry of lens, instead, the lens and screens are taken as a whole with equivalent geometric relationship.

Since the geometric constraints vary within different HMDs, the specific optimal display resolution should be calculated accordingly. HTC VIVE is used in our subjective quality assessment experiment. With specific geometric constraint, 3600×1800 is calculated to be the optimal resolution for HTC VIVE according to the proposed method, which can guarantee a maximized per-pixel display range of more than 20 pixels at the center for each eye.

IV. Subjective Quality Assessment for Panoramic Videos

Other than the presentation problem of the sequences aforementioned, there are also significant differences from that of natural planar video in many aspects during the subjective video quality assessment test. Firstly, the viewing experience during rating test totally differs. On the one hand, most subjects are unfamiliar with the panoramic videos, which emphasizes the importance of a detailed training before formal tests. On the other hand, due to the limitation of developing technologies, the VR viewing with HMDs easily leads to viewing discomfort and thus greatly limits the testing duration compared with the 2D video quality assessment. Furthermore, since the contents of panoramic video exist on every direction and are viewed with VR HMDs as a virtual sphere, it is essential to take the viewing consistency of the subjects into account while conducting the test.

These discrepancies between the natural and panoramic video quality assessment make up an important motivation of our work.To evaluate the quality of panoramic videos, a targeted subjective quality assessment test is to be conducted. This section describes the process with which the subjective quality database for panoramic videos is established.

A. References and Test Sequences

As shown in Fig. 8 and Table I, 10 panoramic common test sequences released by JVET [39] are chosen as reference sequences in our experiment. All the sequences are in the formate of ERP, lasting for 10s each.

Fig. 9 presents the process of test sequence generation. Since the main purpose of the assessment is to evaluate the quality under different coding impairments, resampling the video to a proper resolution before coding can avoid additional quality changes caused by sampling of the HMD. Thus the sequences are first down-sampled to the optimal resolution, i.e., 3600×1800, obtained in Section III with Lanczos sampling method implemented in the 360-Lib Software [41]. Coding impairments are then introduced to the optimally sampled references to obtain test sequences of reconstructed ERP. The reference sequences are compressed using the High Efficiency Video Coding (HEVC) reference software (HM version 16.14) [42] with 360-Lib [41] at 5 quantization parameter values specified in common test conditions [43], i.e., 22, 27, 32, 37, 42, to obtain sequences spanning a relatively wide range of quality in favor of future evaluation and comparison. During the test, Random Access (RA) configuration is used and the IntraPeriod parameters are specified according to 360-Lib, 32 for 30fps and 64 for 60fps. Since HMD cannot support 10-bit video display, the 10-bit sequences are converted to 8 bit with 360-Lib software.

TABLE I
High Fidelity Input Test Sequences in ERP Format [39]

Class	Sequence name	Frame count	Resolution@FPS	Bit-depth
8K	Train_le	600	8192x4096@60	8
8K	SkateboardingTrick_le	600	8192x4096@60	8
8K	SkateboardInLot	300	8192x4096@30	10
8K	ChairLift	300	8192x4096@30	10
8K	KiteFlite	300	8192x4096@30	8
8K	Harbor	300	8192x4096@30	8
4K	PoleVault_le	300	3840x1920@30	8
4K	AerialCity	300	3840x1920@30	8
4K	DrivingInCity	300	3840x1920@30	8
4K	DrivingInCountry	300	3840x1920@30	8

After processing, 5 sequences can be obtained from each reference, which is presented in Fig. 9 as recERP files. Together with the 10 reference sequences, a set of 60 sequences on different but relatively stable quality levels are prepared for the experiment (see Fig. 10 for example).

B. Subjects

30 non-expert subjects are recruited to participate in the assessment experiments. The subjects are undergraduate and graduate students aging from 20 to 26, including 17 males and 13 females. None of the subjects majors in quality assessment, nor do they involve in the design or further analysis of the experiments [21], [48].

All the subjects have normal or corrected-to-normal vision acuity, including far vision, near vision and color vision. Since the VR viewing will easily make people fatigue, those who are severely sick with the VR HMDs are not allowed to participate in the assessment.

C. Experimental Setup

The panoramic videos are displayed with HTC VIVE [49]. Due to the geometrical characteristic mentioned in Section III,

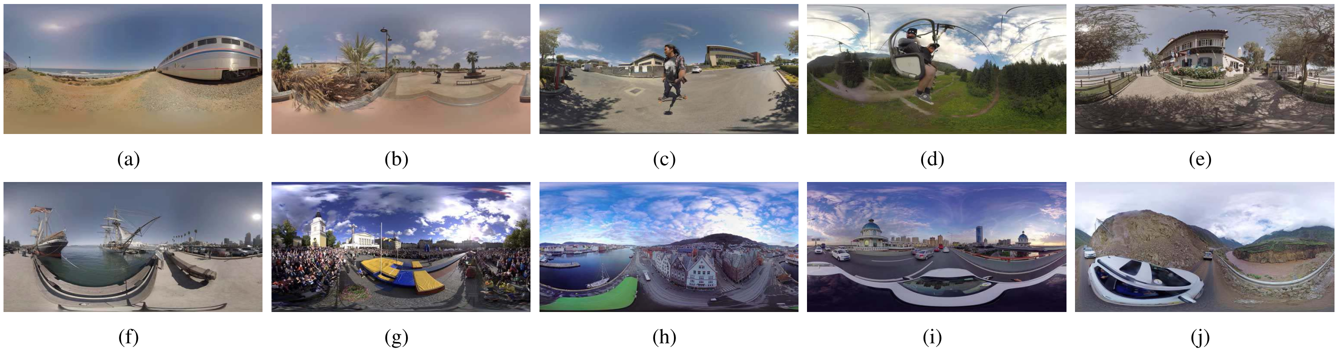


Fig. 8. Example frames of the ten references adopted in the test. (a) Train_le [44], (b) SkateboardingTrick_le [44], (c) SkateboardInLot [44], (d) ChairLift [44], (e) KiteFlite [45], (f) Harbor [45], (g) PoleVault_le [46], (h) AerialCity [47], (i) DrivingInCity [47], (j) DrivingInCountry [47].

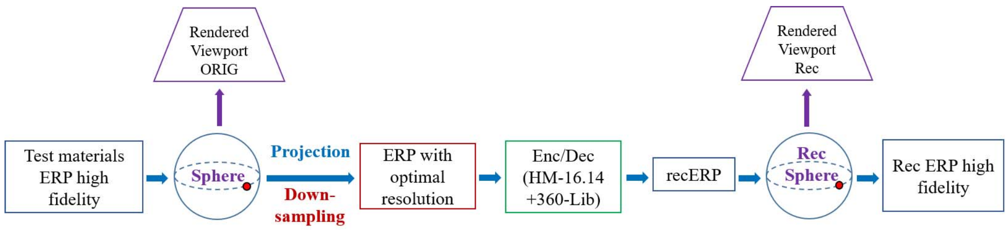


Fig. 9. The process of panoramic test sequence generation [39].

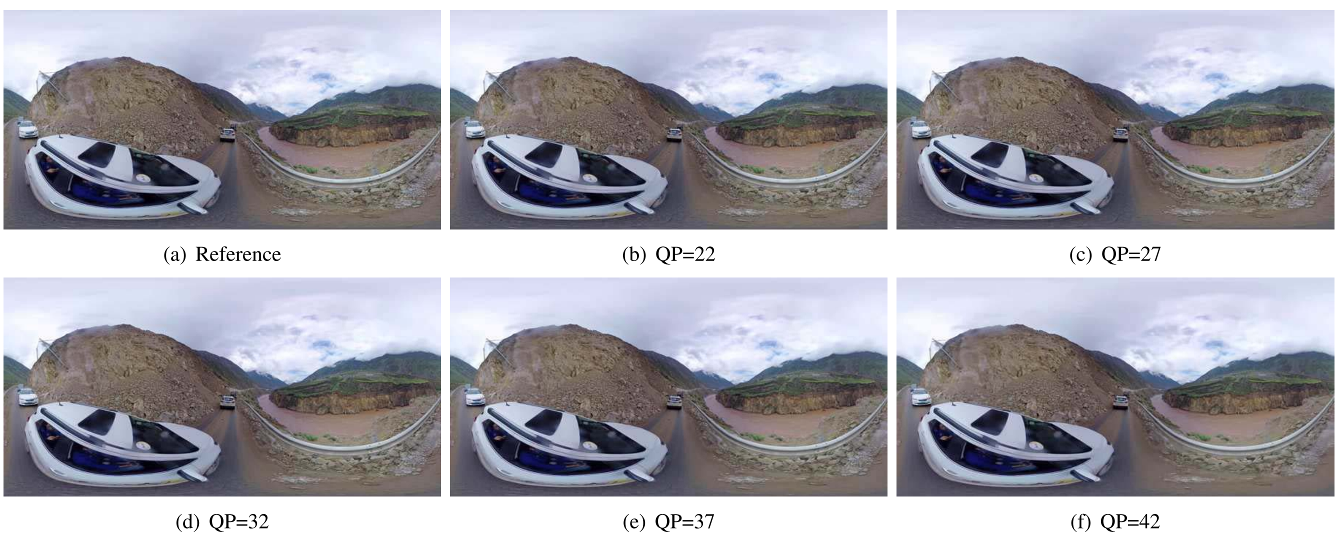


Fig. 10. The reference "DirvingInCountry" and its corresponding impaired sequences at five levels of coding degradation. (a) Reference, (b) QP=22, (c) QP=27, (d) QP=32, (e) QP=37, (f) QP=42.

the original test videos are first sampled to the optimal resolution for the HTC VIVE to guarantee a maximized range of per-pixel display. The sampled videos act as references, and all the impairments being evaluated will be introduced after sampling.

Since panoramic videos shown in the VR HMDs will be displayed in the sphere format, as shown in Fig. 1, the panoramic contents exist in all directions, which leads to the argument whether the contents assessed by different subjects are consistent. To solve the problem, some works like [9] divide the panoramic video into several regions and assess each region stepwise. This solution, to some extent, ignores the important feature of panoramic videos. In order to guarantee a thorough viewing of the omnidirectional content and the real viewing experience, the subjects could move their heads freely to view the video in all directions in our test. The validation of the free-viewing method is discussed in Section V.

D. Assessment Method and Procedure

ACR-HR method is adopted to assess the quality of the videos, which is an effective single stimulus assessment method. The sequences are presented one at a time and are rated independently. The reference sequences will also be presented and rated by the subjects without any special identification, deemed hidden reference. All the test sequences will

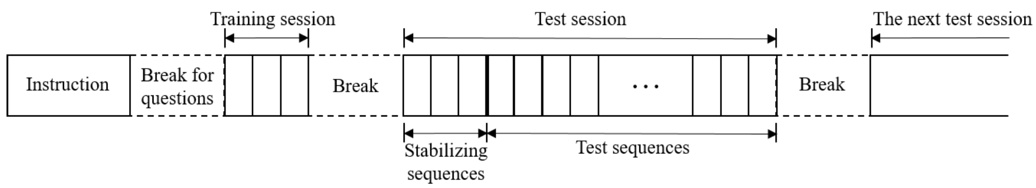


Fig. 11. The subjective assessment procedure.

be presented randomly and each sequence will be displayed only once. The rating scores for the test sequences are defined using the difference between the rating of the test and corresponding reference sequences, i.e., DMOS. To fit for the quality range of the test and make it clearer for non-expert subjects, the five-grade rating scale is used to evaluate the video quality, in which score 1, 2, 3, 4, 5 corresponds to the quality level of "Bad", "Poor", "Fair", "Good", and "Excellent", respectively.

The subjective assessment procedure mainly consists of three phases shown in Fig. 11. The instruction session gives a detailed written instruction on the test to subjects to make them know clearly what to do in different stages during the test, e.g., the aim of the test, the task in each session, the method of assessment, the grading scale to be used, what to evaluate, how and when to vote, number and type of test sequences, total duration of the test, how to deal with discomfort during the test, and thus ensuring a valid process. Since the panoramic video viewing environment differs significantly from that of the planar video, a training session is of much necessity, in which a set of representative panoramic videos ("AerialCity" and its corresponding impaired sequences are used here) are displayed for the subjects to vote so that they can get acquainted with the viewing environment and quality range of the test. The procedure of the test session(s) are shown in Fig. 12. With 3 stabilizing sequences ("ChairLift" and two of its impaired sequences) at the beginning of the first test session and 48 test sequences, the whole viewing process for testing lasts for about 13 minutes, separated into two sessions by a 10-minute rest to avoid viewing fatigue. During the process, the subjects who feel severely uncomfortable can stop at any point of the test, and their data will not be included in the final database.

All the test and reference sequences are presented randomly to avoid order effects. The test sequences and the corresponding reference should not be presented continuously, neither should the test sequences from the same reference. Therefore, a pseudo-random order is used to meet all the conditions [48].

V. Subjective Panoramic Video Quality Assessment Database

After the subjective rating process, each test sequence is assigned 30 rating scores from the subjects participating in the test, based on which the database will be built. In this section, statistical analysis will be conducted to screen and integrate the individual scores for the establishment of subjective quality database.

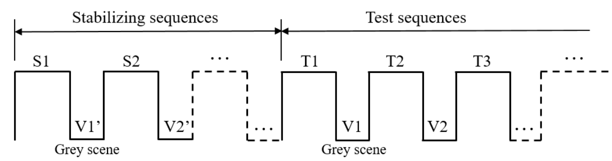


Fig. 12. Structure of a test session. Three stabilizing sequences (S_i) are first presented to stabilize the subjects' rating [21]. Stabilization is only needed in the first session. During the main part of the session, a 10-second test sequence is presented first (T_i), then the subjects vote for the sequence during V_i. Once the score for a sequence is determined, the next 10-second test sequence will then be presented.

A. Subject Reliability

Based on the individual rating data, post-experiment screening is first conducted to assess subject reliability and ensure a valid database. If a subject does not respond according to the instructions, the data have to be discarded. Firstly, a subject will be discarded if there is any missed rating [50]. Secondly, the subject with unreliable ratings will also be screened.

The subject reliability assessment is conducted based on the criteria given in [21]. The Kurtosis of each subject is computed to determine if his/her rating score is normally distributed. If the Kurtosis is between 2 and 4, the subject will be rejected when his/her rating scores on over 5% sequences exceed two standard deviation from the mean score of all the subjects on the corresponding sequences. Otherwise the subject will be discarded when over 5% of his/her scores exceed $\sqrt{20}$ standard deviation from the mean scores.

In total, the ratings from 3 subjects are discarded by the screening process. Therefore, 27 subjects are considered reliable and included in our subjective rating database.

B. Calculation of DMOS

Since the reference sequences are also presented and rated without special identification, DMOS is calculated with the reliable individual ratings as the final scores. First of all, the Differential Viewer scores (DV) are calculated on the basis of hidden reference [20]:

$$DV_{ij} = V_{ij} - V_{ij_{ref}} + 5 \tag{6}$$

where DV_{ij} means the DV of subject i on test sequence j. V_{ij} means the rating score of subject i on sequence j. $V_{ij_{ref}}$ means the rating score of subject i on the reference sequence of test sequence j. During the calculation, any DV greater than 5, i.e., the test sequence is rated better than its reference, will also be accepted. Under this circumstance, a 2-point crushing function

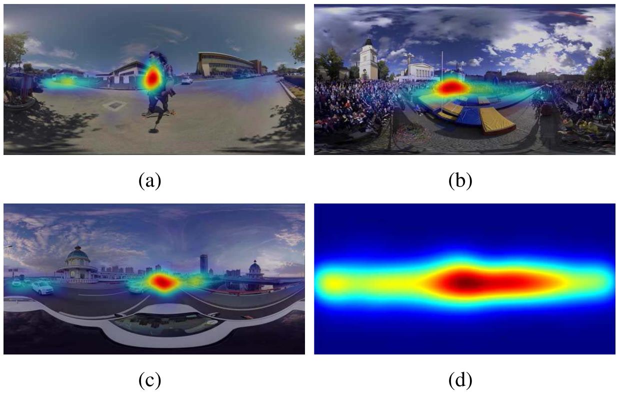


Fig. 13. The heat maps of the head movement on example sequences. (a)-(c) are the heat maps obtained with the head movement data on the corresponding sequence from all the subjects on all the frames and visualized on one of the frames randomly selected from the sequence. (d) shows the heat map combining the head movement data on all the sequences from all the subjects.

specified in [20] will be applied to avoid the influence on the mean opinion score:

$$cDV_{ij} = \frac{7 \times DV_{ij}}{2 + DV_{ij}}, \text{ when } DV_{ij} > 5 \tag{7}$$

Then the *DMOS* of the test sequence j ($DMOS_j$) based on M subjects are calculated as follows:

$$DMOS_j = \frac{\sum_{i=1}^{M} DV_{ij}}{M}. \tag{8}$$

C. Further Validation of the Database

1) Viewing Consistency: Considering the feature of the panoramic video, though the subjects can view freely on 360 degrees, the viewing probability of different areas on the sphere may not be the same. As has been illustrated, the subjects usually pay more attention to the equator, especially the front area of the sphere [10]. Being an important assumption our test protocol based on, the viewing consistency during the rating test is further investigated to validate the proposed method and the corresponding rating results of our test.

During the rating test, the head movement of the subjects are also recorded, which provides evidence for where the subjects pay their attention to during sequence viewing. For each sequence, the head movement data of all the subjects on all the frames are integrated and visualized on one of the frames randomly selected. Some examples are shown in Fig. 13, from which we can see that the viewing direction of different subjects shows a high consistency, i.e., a strong viewing bias towards the equator, especially the front region. On the other hand, Fig. 13(d) combines the viewing direction data from all the subjects on all the sequences into on heat map, which further indicates the consistent viewing tendency.

2) Distribution of the Rating Data: Fig. 14 shows the distribution of the individual rating scores for all the sequences, which exhibits a near-uniform distribution on the whole five-grade quality scale, indicating that the test sequences and coding degradations used in the rating tests are equally distributed and thus may be considered representative.

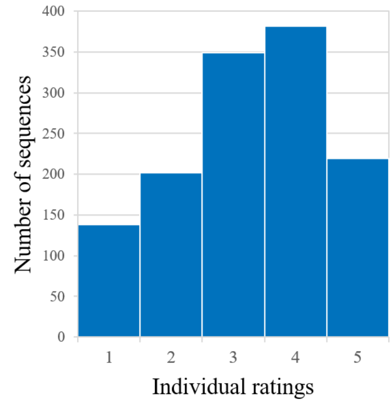


Fig. 14. Distribution of the individual rating scores of all the subjects on all the sequences.

D. Database Summary

As illustrated above, the proposed subjective quality database is considered to be suitable for further applications based on its representative test sequences and reliable subject ratings. Here a summary of the database is made to present the database clearly.

The goal of the database is to investigate the observers' subjective opinion on the quality of panoramic videos with coding artifacts, provide benchmark for objective quality assessment models, and facilitate future coding applications. The database consists of the following contents:

1) Panoramic Video Sequences: 50 distorted panoramic video sequences are generated from 10 8K and 4K raw videos provided by JVET. Using our proposed method in Section III, the references are first down-sampled to an optimal display resolution for HTC VIVE. Coding degradations are then introduced using the HEVC reference software (HM version 16.14) with 360-Lib at 5 quantization parameter values. Among all the sequences, the reference "AerialCity", together with its 5 distorted sequences, is used as training sequences in the training session, while three distorted sequences of "ChairLift" are used as stabilizing sequences at the beginning of the test session. Thus, in total, 40 distorted sequences from the remaining 8 references are used as test sequences in the subjective test.

2) Subjects: 30 non-expert subjects participate in the subjective assessment test, including 17 males and 13 females. The subjects are undergraduate and graduate students with normal or corrected-to-normal vision acuity. They view and evaluate the sequences with HTC VIVE on the video quality. After the post-experiment subject reliability analysis, 27 out of 30 subjects are proved to provide valid subjective rating data for the test sequences in the database.

3) Individual Ratings and DMOS: Since the ratings of training and stabilizing sequences cannot be included in the final results, 1296 individual ratings on the 40 test sequences and 8 references from 27 reliable subjects are finally obtained. The rating is conducted with absolute 5-grade scale on video quality only. As the hidden reference protocol is implemented, DMOS is calculated to represent the quality of the videos.

TABLE II
COMPARISON OF THE CORRELATION BETWEEN OBJECTIVE VIDEO QUALITY ASSESSMENT MODELS AND DMOS. THE MODEL WITH BEST PERFORMANCE IS HIGHLIGHTED IN BOLD

Algorithm	PSNR	WS-PSNR	S-PSNR-NN	**CPP-PSNR**
PLCC	0.7756	0.8028	0.8021	**0.8058**
SROCC	0.7829	0.8022	0.8022	**0.8166**

VI. PERFORMANCE EVALUATION OF OBJECTIVE QUALITY ASSESSMENT METHODS

To evaluate related technologies in a more effective way, there have been some objective quality assessment models that automatically measure the quality of the panoramic videos, among which Peak Signal-to-Noise Ratio (PSNR) is most widely used in measuring coding impairments and is considered as a baseline for the performance of objective assessment methods. Based on PSNR, many metrics specifically designed for panoramic videos, e.g., WS-PSNR [51], S-PSNR [10], are proposed to provide more reliable evaluation results. Despite the convenience of the objective models, subjective quality assessment is the most reliable way to know how the video quality is perceived by human. Therefore, it is essential to evaluate the objective models with subjective quality scores. Further problems can be discovered through the comparison with the subjective quality, and improvements can also be made more efficiently with the help of perceptual attributes. In this section, the performance of several objective quality assessment models adopted by JVET is evaluated with the proposed subjective quality database.

The following models are evaluated in this section[1]:

- *Peak Signal-to-Noise Ratio (PSNR):* Calculates PSNR based on all samples with equal weight.
- *Weighted to Spherically uniform PSNR (WS-PSNR) [51]:* Evaluates difference in spherical domain by weighting each pixel with its area proportion on the sphere.
- *Sphere based PSNR at Nearest Neighbor (S-PSNR-NN) [39]:* S-PSNR calculates PSNR based on the points uniformly sampled on the sphere surface. S-PSNR-NN is one of S-PSNR's variants that evaluates the distortion at nearest neighbor integer sample positions, rather than at the fractional sampling positions to avoid the influence of interpolation.
- *PSNR for Carster Parabolic Projection (CPP-PSNR) [52]:* Compares quality across different projection methods using equal area projection (Carster Parabolic Projection). Allows the quality comparison between the sequences of different projection schemes.

Two metrics are used for evaluating the performance of the aforementioned objective assessment methods, i.e., the Spearman's Rank Order Correlation Coefficient (SROCC) and the Pearson Linear Correlation Coefficient (PLCC). SROCC assesses monotonic relationships between objective scores and subjective opinions. The Spearman correlation will be high when a similar rank is observed between the two variables.

[1]The assessment scores of the objective models being evaluated in this paper is measured on Y channel, except for PSNR, which is calculated with YUV channels as the 6:1:1 weighted sum of each channel.

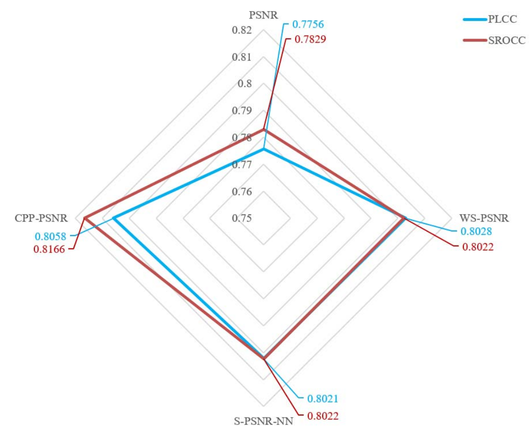


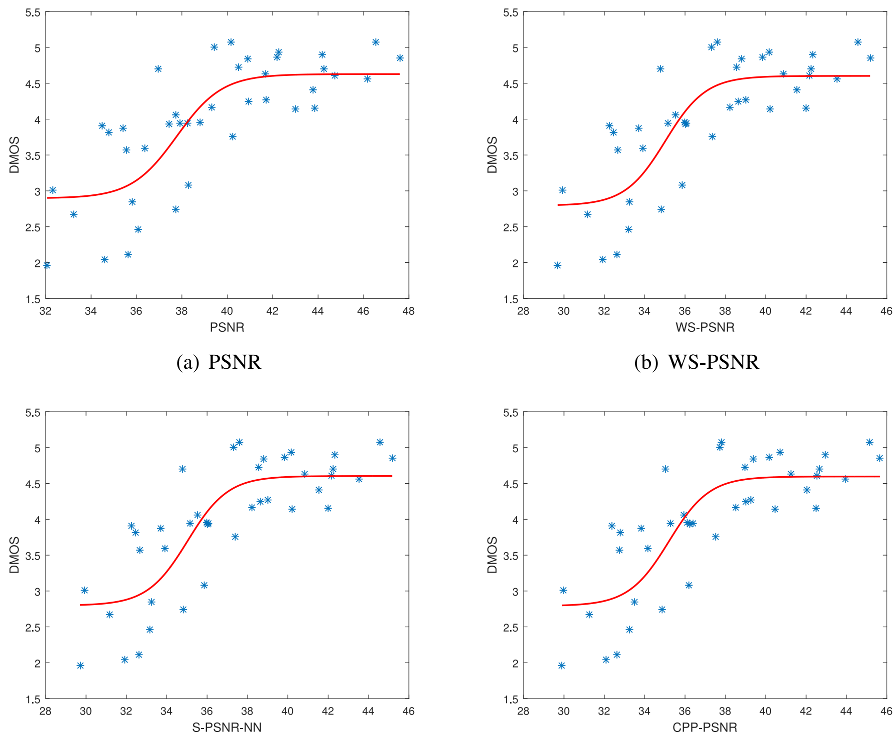
Fig. 15. Comparison of the correlation between objective video quality assessment models and DMOS.

PLCC measures the linear correlation between two variables, which has a value between ± 1, corresponding to the linear correlation from total positive to total negative. To calculate SROCC and PLCC with the DMOS, a nonlinear regression is first performed on the objective scores using a logistic function defined in [50] as follows to fit the scores to the DMOS:

$$Q'_j = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-(Q_j - \beta_3/|\beta_4|)}} \tag{9}$$

where Q_j shows the objective score for sequence j and Q'_j shows the fitted score. The initial value of parameters $\beta_i (i = 1, 2, 3, 4)$ are determined according to [50] and optimized with nonlinear least squares optimization.

Table II and Fig. 15 shows the performance of correlation analysis. Scatter plots of the objective scores and DMOS on the 48 test sequences are shown in Fig. 16, in which the red line represents the best fitting logistic curve. Since the newly emerging panoramic videos show much less variation on their compressing type, content, scene, texture or motion compared with the natural videos, even the conventional PSNR achieves a moderate performance comparing with that on the natural ones. The similar phenomenon is also presented in some researches on quality assessment of 3D videos and images [53], [54]. Especially in [54], which presents the performance of PSNR on two 3D datasets of different sizes. PSNR also performs much better on the smaller dataset than that on the natural sequences and the larger dataset. However, PSNR is calculated based on all samples with equal weight, ignoring the spherical feature of panoramic videos. The models being optimized with the characteristic of sphere achieves improvements compared with the simple PSNR model and are also theoretically more reasonable, which demonstrates the necessity of modeling the features of spherical domain such as



(a) PSNR (b) WS-PSNR

(c) S-PSNR-NN (d) CPP-PSNR

Fig. 16. Scatter plots of objective video quality assessment models and DMOS for all the test sequences. The red line in each plot shows the best fitting logistic curve. (a) PSNR, (b) WS-PSNR, (c) S-PSNR-NN, (d) CPP-PSNR.

projection, pixel weight, etc. Among all the models being evaluated, CPP-PSNR performs best in terms of both SROCC and PLCC, from which the importance of excluding the interference brought by projection while evaluating the quality with respect to coding impairments can be verified. Furthermore, the comparison also shows that the performance promotion is limited. The objective quality assessment of panoramic videos still awaits further researches and improvements.

VII. Conclusion

In this paper, a subjective panoramic video quality assessment protocol is proposed for coding applications, which takes the features of panoramic videos into account. Considering the projection and the resolution limitation of HMDs, the method of re-sampling the video sequence to an optimal resolution before coding is proposed first. With the optimal display resolution, a maximized range of per-pixel display on the center area of the video can be guaranteed, alleviating unexpected quality change caused by sampling of the HMDs and thus making the assessment more reliable. Furthermore, a subjective quality database is established based on the proposed protocol, including 50 distorted sequences generated from 10 raw panoramic videos using HEVC compression, subjective rating scores from 27 reliable subjects and DMOSs of the test sequences. Based on the proposed database, the performance of several existing objective quality assessment models suggested by JVET is evaluated, which shows the superiority over the conventional baseline and, meanwhile, indicates the need for further researches and improvements.

Acknowledgment

The authors would like to thank VQEG members for the helpful discussions and Immersive Media Group for the great support.

References

[1] A. Wexelblat, *Virtual Reality: Applications and Explorations*. San Diego, CA, USA: Academic Press, 2014.

[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,” *Front. Psychol.*, vol. 6, no. 26, pp. 1–9, 2015.

[3] 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.

[4] J. D. N. Dionisio, W. G. Burns, III, and R. Gilbert, “3D virtual worlds and the metaverse: Current status and future possibilities,” *ACM Comput. Surveys*, vol. 45, no. 3, pp. 1–38, 2013.

[5] K.-T. Ng, S.-C. Chan, and H.-Y. Shum, “Data compression and transmission aspects of panoramic videos,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 82–95, Jan. 2005.

[6] W. Zhou, W. Qiu, and M.-W. Wu, “Utilizing dictionary learning and machine learning for blind quality assessment of 3-D images,” *IEEE Trans. Broadcast.*, vol. 63, no. 2, pp. 404–415, Jun. 2017.

[7] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, “Objective video quality assessment methods: A classification, review, and performance comparison,” *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[8] L. Lu, Z. Wang, A. C. Bovik, and J. Kouloheris, “Full-reference video quality assessment considering structural distortion and no-reference quality evaluation of MPEG video,” in *Proc. IEEE Int. Conf. Multimedia Expo*, Lausanne, Switzerland, 2002, pp. 61–64.

[9] V. Zakharchenko, K. P. Choi, and J. H. Park, “Quality metric for spherical panoramic video,” in *Proc. SPIE Opt. Photon. Inf. Process. X*, vol. 9970. San Diego, CA, USA, 2016, pp. 1–9.

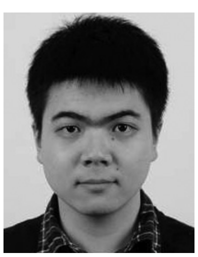
[10] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Fukuoka, Japan, 2015, pp. 31–36.
[11] S. Leorin, L. Lucchese, and R. G. Cutler, "Quality assessment of panorama video for videoconferencing applications," in *Proc. IEEE 7th Workshop Multimedia Signal Process.*, Shanghai, China, 2005, pp. 1–4.
[12] *ITU-R Recommendations*. Accessed: Oct. 2017. [Online]. Available: http://www.itu.int/rec/R-REC-BT/
[13] *Video Quality Experts Group*. Accessed: Jun. 2017. [Online]. Available: https://www.its.bldrdoc.gov/vqeg
[14] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.
[15] M. Urvoy *et al.*, "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," in *Proc. 4th Int. Workshop Qual. Multimedia Exp.*, 2012, pp. 109–114.
[16] L. Goldmann, F. D. Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," in *Proc. SPIE Three Dimensional Image Process. Appl.*, vol. 7526. San Jose, CA, USA, 2010, pp. 158–160.
[17] A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. de Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 652–671, Oct. 2012. [Online]. Available: http://live.ece.utexas.edu/research/quality/live_mobile_video.html
[18] J. Lee *et al.*, "A stereoscopic 3-D broadcasting system using fixed and mobile hybrid delivery and the quality assessment of the mixed resolution stereoscopic video," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 222–237, Jun. 2015.
[19] 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.
[20] "Subjective video quality assessment methods for multimedia applications," ITU-T, Geneva, Switzerland, ITU-T Recommendation P. 910, 2008.
[21] "Methodology for the subjective assessment of the quality of television pictures," ITU-R, Geneva, Switzerland, ITU-T Recommendation BT. 500-13, 2012.
[22] "Subjective audiovisual quality assessment methods for multimedia applications," ITU-T, Geneva, Switzerland, ITU-T Recommendation P. 911, 1998.
[23] M. H. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," in *Proc. SPIE Vis. Commun. Image Process.*, vol. 5150, 2003, pp. 573–582.
[24] S. Péchard, R. Pépion, and P. Le Callet, "Suitable methodology in subjective video quality assessment: A resolution dependent paradigm," in *Proc. Int. Workshop Image Media Qual. Appl.*, 2008, p. 6.
[25] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik. (2016). *LIVE Netflix Video Quality of Experience Database*. [Online]. Available: http://live.ece.utexas.edu/research/LIVE_NFLXStudy/index.html
[26] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 626–639, Oct. 2012.
[27] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process. Image Commun.*, vol. 30, pp. 57–77, Jan. 2015.
[28] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4408–4421, Nov. 2015.
[29] J. Boyce and Z. Deng, "AHG8: Subjective testing of 360° video projection/packing formats," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-F0021, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2017.
[30] Z. Deng, L. Xu, and J. Boyce, "AHG8: Subjective test pilot study of 360° video projection/packing formats," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-F0083, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2017.
[31] K. Kawamura and S. Naito, "Comments on subjective testing procedure of 360° video," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-F0067, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2017.
[32] V. R. Gaddam, M. Riegler, R. Eg, C. Griwodz, and P. Halvorsen, "Tiling in interactive panoramic video: Approaches and evaluation," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1819–1831, Sep. 2016.
[33] Y. Zhang *et al.*, *Test Plan for Subjective Assessment of VR Video Quality*, VQEG Immersive Media Group, London, U.K., 2016. [Online]. Available: ftp://vqeg.its.bldrdoc.gov/Documents/VQEG_London_Oct16/
[34] Z. Chen and Y. Zhang, "Towards subjective quality assessment for panoramic video," in *Proc. SPIE Human Vis. Electron. Imag.*, 2018, pp. 1–5.
[35] A. Singla, S. Fremerey, W. Robitza, and A. Raake, "Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays," in *Proc. Int. Conf. Qual. Multimedia Exp.*, Erfurt, Germany, 2017, pp. 1–6.
[36] P. Pérez, *Project Vertigo: Monitoring Sickness and Discomfort in High-Motion 360 Video*, VQEG Immersive Media Group, Los Gatos, CA, USA, 2017. [Online]. Available: ftp://vqeg.its.bldrdoc.gov/Documents/VQEG_Los_Gatos_May17/
[37] M. Urvoy, M. Barkowsky, and P. Le Callet, "How visual fatigue and discomfort impact 3D-TV quality of experience: A comprehensive review of technological, psychophysical, and psychological factors," *Ann. Telecommun.*, vol. 68, nos. 11–12, pp. 641–655, 2013.
[38] M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx, "Visual discomfort of 3D TV: Assessment methods and modeling," *Displays*, vol. 32, no. 4, pp. 209–218, 2011.
[39] J. Boyce, E. Alshina, A. Abbas, and Y. Ye, "JVET common test conditions and evaluation procedures for 360° video," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-D1030, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2017.
[40] V. Zakharchenko, E. Alshina, K. P. Choi, M. Choi, A. Dsouza, and A. Singh, "AHG8: I cosahedral projection for 360-degree video content," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-D0028, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2016.
[41] Y. He *et al.*, "JVET 360Lib software manual," document ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2016. [Online]. Available: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/trunk/
[42] F. Bossen, D. Flynn, K. Sharman, and K. Sühring, "HM 16.14 software manual," document ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Joint Collaborative Team Video Coding, ITU-T, Geneva, Switzerland, 2016. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.14/doc/software-manual.pdf
[43] K. Sühring and X. Li, "JVET common test conditions and software reference configurations," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-B1010, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2016.
[44] A. Abbas and B. Adsumilli, "AHG8: New GoPro test sequences for virtual reality video coding," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-D0026, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2016.
[45] E. Asbun, Y. He, Y. He, and Y. Ye, "AHG8: InterDigital test sequences for virtual reality video coding," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-D0039, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2016.
[46] S. Schwarz *et al.*, "Tampere pole vaulting sequence for virtual reality video coding," document ITU-T SG16 WP3M, ISO/IEC JTC1/SC29/WG11, and JVET-D0143, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2016.
[47] W. Sun and R. Guo, "Test sequences for virtual reality video coding from LetinVR," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-D0179, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2016.
[48] VQEG 3DTV Group. (2012). *Test Plan for Evaluation of Video Quality Models for Use With Stereoscopic Three-Dimensional Television Content*. [Online]. Available: https://www.its.bldrdoc.gov/vqeg/projects/3dtv/3dtv.aspx
[49] *HTC VIVE*. Accessed: Jun. 2017. [Online]. Available: https://www.vive.com/
[50] VQEG FRTV Phase I. (2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. [Online]. Available: https://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-i/frtv-phase-i.aspx
[51] Y. Sun, A. Lu, and L. Yu, "AHG8: WS-PSNR for 360 video objective quality evaluation," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-D0040, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2016.
[52] V. Zakharchenko, E. Alshina, K. P. Choi, A. Singh, and A. Dsouza, "AHG8: Suggested testing procedure for 360-degree video," document ITU-T SG16 WP3, ISO/IEC JTC1/SC29/WG11, and JVET-D0027, Joint Video Explor. Team, ITU-T, Geneva, Switzerland, 2016.

[53] A. Banitalebi-Dehkordi, M. T. Pourazad, and P. Nasiopoulos, “3D video quality metric for 3D video compression,” in *Proc. IVMSP*, Seoul, South Korea, 2013, pp. 1–4.
[54] J. Kim, T. Kim, S. Lee, and A. C. Bovik, “Quality assessment of perceptual crosstalk on two-view auto-stereoscopic displays,” *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4885–4899, Oct. 2017.

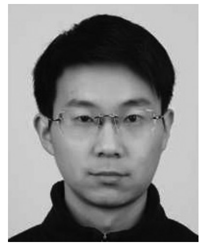
Yingxue Zhang received the B.Eng. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2015, where she is currently pursuing the Ph.D. degree with the School of Remote Sensing and Information Engineering. Her research interests include quality assessment, computational vision, and image and video processing. She was a recipient of National Scholarship in 2015.

Yingbin Wang received the B.Eng. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2016, where he is currently pursuing the master’s degree with the School of Remote Sensing and Information Engineering. His research interests include multimedia technology, HCI, and image and video processing.

Feiyang Liu received the B.Eng. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2016, where he is currently pursuing the Ph.D. degree with the School of Remote Sensing and Information Engineering. His research interests include image and video processing, data compression, and deep learning.

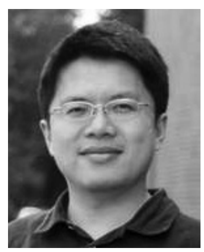


Zizheng Liu received the B.Eng. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2016, where he is currently pursuing the Ph.D. degree with the School of Remote Sensing and Information Engineering. His research interests include image and video processing, multimedia technology, and deep learning.



Yiming Li received the B.Eng. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2015, where he is currently pursuing the Ph.D. degree with the School of Computer Science. His research interests include image and video processing, and multimedia technology. He was a recipient of National Scholarship in 2016/2017.

Daiqin Yang (M’13) received the B.E. and M.E. degrees from the Huazhong University of Science and Technology, China, and the Ph.D. degree from the University of Hong Kong, all in electrical and electronic engineering. She was with Philips Research Asia, Shanghai, Nanyang Technological University, Singapore, and the Institute for Infocomm Research, Singapore. She joined the faculty of Wuhan University, China, where she is currently an Associate Professor with the School of Remote Sensing and Information Engineering. Her research interests include sensor networks, intelligent systems, and remote sensing image processing.



Zhenzhong Chen (S’02–M’07–SM’15) received the B.Eng. degree from the Huazhong University of Science and Technology and the Ph.D. degree from the Chinese University of Hong Kong, both in electrical engineering. He was with MediaTek USA Inc., San Jose, CA, USA. He is currently a Professor with Wuhan University. His current research interests include image processing and understanding, computer vision, HCI, multimedia communications, photogrammetry, and remote sensing.

He was a recipient of the CUHK Young Scholar Dissertation Award, the CUHK Faculty of Engineering Outstanding Ph.D. Thesis Award, Microsoft Fellowship, ERCIM Alain Bensoussan Fellowship, First Class Prize of 2015 IEEE BigMM Challenge, and IEEE SMC 2015 Best Paper Franklin V. Taylor Memorial Award Finalist. He has been selected for the Thousand Talents Plan for Young Professionals. He has been a VQEG Board Member and the Immersive Media Working Group Co-Chair, a Selection Committee Member of ITU Young Innovators Challenges, a member of the IEEE Multimedia Systems and Applications Technical Committee, and the Co-Chair of IEEE Multimedia Communication TC Networking Technologies for Multimedia Communication IG. He is an Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *Journal of the Association for Information Science and Technology*, and the *Journal of Visual Communication and Image Representation* and an Editor of IEEE IoT NEWSLETTER. He has been the Special Session Chair of IEEE World Forum of Internet of Things 2014, the Publication Chair of IEEE Conference on Multimedia and Expo 2014, the Special Session Chair of VCIP 2016, and has served as an Area Chair for ICME and VCIP.