

Visual Quality Assessment of Panoramic Video

Mai Xu, *Senior Member, IEEE*, Chen Li, *Student Member, IEEE*, Zulin Wang, *Member, IEEE*,
and Zhenzhong Chen *Senior Member, IEEE*

Abstract—In contrast with traditional video, panoramic video enables spherical viewing direction with support for head-mounted displays, providing an interactive and immersive experience. Unfortunately, to the best of our knowledge, there are few visual quality assessment (VQA) methods, either subjective or objective, for panoramic video. This paper proposes both subjective and objective methods for assessing quality loss in impaired panoramic video. Specifically, we first establish a new database, which includes the viewing direction data from several subjects watching panoramic video sequences. Then, from our database, we find a high consistency in viewing direction across different subjects. The viewing directions are normally distributed in the center of the front regions, but they sometimes fall into other regions, related to video content. Given this finding, we present a subjective VQA method for measuring different mean opinion score (DMOS) of the whole and regional panoramic video, in terms of overall DMOS (O-DMOS) and vectorized DMOS (V-DMOS), respectively. Moreover, we propose two objective VQA methods for panoramic video, in light of human perception characteristics of panoramic video. One method weighs the distortion of pixels with regard to their distances to the center of front regions, which considers human preference in a panorama. The other method predicts viewing directions according to video content, and then the predicted viewing directions are leveraged to assign weights to the distortion of each pixel in our objective VQA method. Finally, our experimental results verify that both the subjective and objective methods proposed in this paper advance state-of-the-art VQA for panoramic video.

Index Terms—Panoramic video, visual quality assessment (VQA), viewing direction

I. INTRODUCTION

Recent years have witnessed the rapid development of virtual reality (VR). According to a report by [1], 90% VR content is in the form of panoramic video, which involves a $360^\circ \times 180^\circ$ viewing direction. With the support of head-mounted displays (HMD), panoramic video offers an immersive and even interactive visual experience [2]. On the other hand, it is likely that the quality of experience (QoE) [3] of panoramic video dramatically degrades when presented at low resolutions or with compression artifacts. Such QoE degradation always makes humans feel uncomfortable, as reported in the MPEG survey [4]. Therefore, it is necessary to study on visual quality assessment (VQA) for panoramic video.

Both subjective and objective methods are needed for a Full-Reference (FR) VQA on panoramic video. Subjective VQA

refers to measuring the quality of panoramic video as rated by humans. Since panoramic video ultimately outputs to human eyes, subjective VQA is more rational than objective VQA in assessing visual quality. In addition, subjective VQA can be also used to verify the effectiveness of objective VQA. Unfortunately, to the best of our knowledge, there are few subjective VQA methods [5], [6] for measuring the quality reduction of panoramic videos. In contrast, there are a great number of subjective VQA methods for traditional 2D videos, such as [7]–[9]. In these methods, the mean opinion score (MOS) [10] and differential MOS (DMOS) [11] are widely used metrics for subjective VQA. In this paper, we thus propose a subjective VQA method to assess quality loss in impaired panoramic video in the form of DMOS.

For objective VQA, the spherical characteristic of panoramic video has been taken into account in the latest work of [6], [12]. For example, Yu *et al.* [12] proposed a sphere-based peak signal-to-noise ratio (S-PSNR), which calculates PSNR based on a set of uniformly sampled points on sphere instead of rectangularly mapped pixels. By applying interpolation algorithms, S-PSNR is able to cope with objective quality assessment for panoramic video under different projections. The main difference between 2D and panoramic videos is that only content inside the field of view (FoV) is accessible in panoramic video. However, none of the existing VQA methods takes into consideration such perceptual characteristics of panoramic video. In this paper, we further propose to objectively assess the perceptual quality of panoramic video by considering the accessible FoV of possible viewing directions.

To be more specific, this paper first establishes a new database containing the viewing directions of 40 subjects on watching 48 panoramic video sequences. Then, by mining our database, we discover that the viewing directions across different subjects are highly consistent. In light of such a finding, we develop two subjective VQA metrics, namely the overall DMOS (O-DMOS) and vectorized DMOS (V-DMOS), for rating the overall and regional visual quality reduction of panoramic video, respectively. We further find from our database that the consistent viewing directions of humans are related to both panoramic location and video content. Accordingly, we propose two objective VQA methods that assign weights to the distortion of each pixel when calculating the PSNR. The weight assignment in the first method only leverages humans' preferences for the location of pixels in panoramic video, while the second method also depends on video content.

This paper is an extended version of our conference paper

M. Xu, C. Li and Z. Wang are with the School of Electronic and Information Engineering, Beihang University, Beijing, 100191 China (e-mail: Maixu@buaa.edu.cn; jnlichen123@buaa.edu.cn; wzulin@buaa.edu.cn). Z. Chen is with Wuhan University, Wuhan, China (e-mail: zzchen@whu.edu.cn). This work was supported by NSFC under grant number 61573037.

TABLE I
PANORAMIC VIDEO TEST SEQUENCE CATEGORIES.

Category	Computer Animation (CA)	Driving	Action Sports	Movie	Video Game	Scenery	Show	Others	In Total
Number of Video sequences	6	6	6	6	6	6	6	6	48

[13]. Beyond the subjective VQA method in [13], this paper further proposes two objective VQA methods for measuring the quality of panoramic video. For the objective VQA methods, this paper also investigates some new findings about the distribution of human viewing direction in panoramic video. Our contributions in this paper are three-fold:

- We establish a viewing direction database for panoramic video, with a consistency analysis on the viewing directions of different subjects. To our best knowledge, our database is the first one collecting viewing direction data of panoramic video.
- We develop a new method for the subjective VQA of panoramic video, taking advantage of the analysis of viewing directions over our database. Our subjective VQA method has been adopted [14] by the international standard IEEE 1857.9/AVS-VR.
- We propose two methods for the objective VQA of panoramic video, taking into account human perception related to panoramic location and video content, respectively. Our objective VQA methods are pioneering work that embeds human perception in assessing visual quality of panoramic video.

II. RELATED WORK

A. Related work on subjective VQA

The past two decades have witnessed a number of subjective VQA methods for 2D video. In particular, the international telecommunication union (ITU) has proposed several subjective methodologies [7]–[9] for assessing 2D video. Among these proposals, the double stimulus continuous quality scale (DSCQS) [15], single stimulus continuous quality scale (SS-CQS) [11] and single stimulus continuous quality evaluation (SSCQE) [16] were adopted to determine the display orders of sequences when viewing and rating video sequences. Additionally, two metrics have been widely used in rating the subjective VQA of 2D video: one metric is MOS [10] for no-reference (NR), reduced-reference (RR) and FR assessments; the other metric is DMOS [11], [17], which is for FR assessment only. Recently, several subjective VQA methods for other types of videos have emerged. For example, Pourashraf *et al.* [18] proposed measuring the subjective quality of video conferencing by adopting DMOS for the conventional subjective VQA method. ITU extended their DMOS-based VQA method for stereoscopic video [19], which incorporates the characteristics of stereoscopic video.

Although panoramic video is flooding into our daily life, there are few works in the literature [5], [6] on the subjective VQA of panoramic video. Upenik *et al.* [5] proposed a testbed for subjective VQA on panoramic video and image. In their testbed, an HMD is suggested as the displaying device, and

a custom software application is provided. Unfortunately, [5] does not deal with how to measure the subjective quality of panoramic video. To the best of our knowledge, the only work on measuring the subjective quality of panoramic video was presented in [6], in which subjects were forced to view one region of panoramic video, and then the conventional subjective VQA method for 2D video is simply applied. However, this is not in accordance with the interactive experience on panoramic video. More importantly, an immersive experience cannot be achieved in the subjective VQA of [6], such that the resulting DMOS does not meet practical QoE for humans. In this paper, we propose a subjective VQA method that considers the interactive behavior of humans in viewing panoramic video, such that the QoE of subjects can be reflected in our subjective metric.

B. Related work on objective VQA

For objective VQA of 2D video, a commonly used FR metric is PSNR. PSNR is based on the mean squared errors (MSE) between the reference and processed videos and has been well studied from a mathematical perspective. However, PSNR cannot successfully reflect the subjective visual quality perceived by the human visual system (HVS), as it does not consider human perception at all. For example, the subjective quality is more likely to be influenced by the PSNR in ROI regions. In order to better correlate assessments with subjective quality, many advanced PSNR-based methods [20]–[26] have been proposed to improve the existing PSNR metric for the VQA of 2D video by accounting for the importance of each pixel. For example, based on the foveation response of HVS, foveal PSNR (FPSNR) [20] was proposed, using a non-uniform resolution weighting metric, in which the distortion weights decrease with eccentricity. The peak signal-to-perceptible noise ratio (PSPNR) [21] and foveated PSPNR [24] were presented to consider distortion, only when the errors are larger than the just-noticeable-distortion (JND) thresholds. Similarly, semantic PSNR (SPSNR) [22] and eye-tracking-weighted PSNR (EWPSNR) [23] were developed based on the form of PSNR as well. EWPSNR has a better performance in evaluating visual quality according to the real-time detected eye fixation points. In [25], a weight-based PSNR metric was proposed to measure the quality of video conferencing. Their method imposes a greater penalty weight on regions with faces and facial features when calculating the PSNR. Free energy adjusted PSNR (FEA-PSNR) was proposed in [26]. This method considers image perceptual complexity when assessing image quality. In [27], a non-reference PSNR method was proposed to assess both quantization error and the blocky effect in measuring the non-reference quality of H.264/AVC videos.



Fig. 1. Heat maps of viewing directions on some selected sequences. Note that the heat maps are obtained via the Gaussian convolution of videoing direction data for all frames viewed by 40 subjects, and the results are shown together with one randomly selected frame from each sequence.

For panoramic video, there are several objective VQA works [6], [12], [28], also based on PSNR. In evaluating the quality degradation of panoramic video encoding, the work of [6], [12] takes into account the spherical characteristic of panoramic video. For example, Yu *et al.* [12] proposed a sphere-based PSNR (S-PSNR), which calculates PSNR based on a set of uniformly sampled points on a sphere instead of rectangularly mapped pixels. By applying interpolation algorithms, S-PSNR is able to generate objective quality assessments for panoramic videos under different projections. Besides, Zakharchenko *et al.* [6] proposed a weighted PSNR (W-PSNR) using gamma-corrected pixel values for the PSNR calculation process. The latest work of [28] conducted an experiment to evaluate the performance of several objective VQA methods on panoramic images, via measuring the correlation between objective and subjective quality. The experimental results reveal that the VQA methods designed for panoramic content slightly outperform traditional VQA methods for 2D content. This finding is probably because none of the existing VQA methods explores the human perception model for panoramic video, in which only content inside FoV is accessible. Therefore, this paper further proposes to objectively assess visual quality of panoramic video, by considering the FoV of possible viewing directions.

III. ANALYSIS OF CONSISTENCY IN VIEWING PANORAMIC VIDEO

Due to the omnidirectionality of panoramic video, people cannot see the whole video at one sight. Instead, they normally look around and focus on what attracts them. It is intuitive that there may exist consistency across different subjects in their viewing directions on watching panoramic video. Thus, this section mainly discusses the analysis of consistency in viewing panoramic video.

A. Database

We establish a new database that contains viewing direction data from 40 subjects when watching panoramic video sequences. In all, there are 48 sequences of panoramic video in our database. To ensure QoE, the resolution of the sequences is beyond 3K (2880×1440) and up to 8K (7680×3840). These sequences are diverse in terms of their content, and they can be categorized according to video content, as shown in Table I. All of these 48 sequences were downloaded from YouTube or VRcun. Then, the sequences were cut into short clips with durations ranging from 20 to 60 seconds. The audio tracks were discarded to avoid the impacts of acoustic information.

We used the HTC Vive as the HMD and a software Virtual Desktop (VD) as the panoramic video player. In total, 40 subjects (29 males and 11 females) participated in the experiment. For each subject, all of the 48 sequences were played in a random order. During the experiment, the subjects were seated in a swivel chair and were allowed to turn around freely, such that all regions of panoramic video were accessible. Besides, to avoid eye fatigue and motion sickness, there was a 5-minute interval between every 16-sequence session. With the support of the Vive software development kit (SDK), we were able to collect the posture data of subjects when viewing panoramic video. Then, the viewing direction data describing where subjects paid attention were obtained in the form of Euler angles, and only the inclination and azimuth angles were recorded. Based on the inclination and azimuth angles, viewing directions of each subject, in terms of longitude and latitude, were collected for the panoramic video sequences in our database. Our database is available at <https://github.com/Archer-Tatsu/head-tracking>.

B. Data analysis

We now analyze the viewing direction data in our database. First, we discard the viewing direction data of the first second in each sequence since the viewing directions of all subjects were initialized to be in the center of the front region. The remaining data are then used for our analysis. Our findings with the corresponding analysis are presented and analyzed as follows.

Finding 1: When subjects are watching panoramic video, the longitude and latitude of their viewing directions are almost uncorrelated with each other.

The viewing direction data we collected in Section III-A consist of two dimensions, i.e., the longitude and latitude in a spherical coordinate system. Let φ and θ denote the collections of the longitude and latitude of viewing directions, respectively, from all panoramic video sequences in our database. Then, the covariance between φ and θ can be calculated as follows,

$$\text{cov}(\varphi, \theta) = E[(\varphi - E(\varphi))(\theta - E(\theta))]. \quad (1)$$

Given the covariance of (1), the correlation between the longitude and latitude of viewing direction can be computed by

$$\rho(\varphi, \theta) = \frac{\text{cov}(\varphi, \theta)}{\sqrt{\text{var}(\varphi)}\sqrt{\text{var}(\theta)}}, \quad (2)$$

where $\text{var}(\varphi)$ and $\text{var}(\theta)$ are the variances of φ and θ , respectively. In our database, we have $\rho(\varphi, \theta) = -0.0337$

TABLE II
CC OF VIEWING DIRECTION HEAT MAPS BETWEEN GROUPS *A* AND *B* FOR EACH PANORAMIC VIDEO SEQUENCE

Category	Name	CC	Category	Name	CC	Category	Name	CC	Category	Name	CC
CA	AcerPredator	0.839±0.087	Driving	AirShow	0.783±0.078	Others	A380	0.839±0.106	Video Game	CS	0.819±0.084
	BFG	0.644±0.146		DrivingInAlps	0.857±0.071		CandyCarnival	0.723±0.094		Dota2	0.714±0.103
	CMLauncher	0.828±0.119		F5Fighter	0.592±0.126		MercedesBenz	0.592±0.133		GalaxyOnFire	0.762±0.084
	Cryogenian	0.526±0.174		HondaF1	0.872±0.053		RingMan	0.897±0.054		LOL	0.724±0.097
	LoopUniverse	0.779±0.078		Rally	0.867±0.047		RioOlympics	0.624±0.123		MC	0.726±0.115
	Pokemon	0.607±0.182		Supercar	0.854±0.064		VRBasketball	0.770±0.105		SuperMario64	0.860±0.054
Movie	Help	0.859±0.122	Scenery	Antarctic	0.674±0.135	Show	BTSRun	0.867±0.061	Action Sports	Gliding	0.528±0.158
	IRobot	0.771±0.078		BlueWorld	0.559±0.156		Graffiti	0.807±0.100		Parachuting	0.628±0.157
	Predator	0.696±0.124		Dubai	0.646±0.133		KasabianLive	0.722±0.132		RollerCoaster	0.834±0.078
	ProjectSoul	0.918±0.053		Egypt	0.665±0.131		NotBeAloneTonight	0.587±0.131		Skiing	0.766±0.104
	StarWars	0.950±0.016		StarryPolar	0.495±0.152		Symphony	0.779±0.096		Surfing	0.830±0.096
	Terminator	0.843±0.078		WesternSichuan	0.667±0.138		VRBasketball	0.770±0.105		Waterskiing	0.781±0.128
Overall		0.745± 0.114									

for the viewing directions of all panoramic video sequences. Since such a value of $\rho(\varphi, \theta)$ is approximately equivalent to 0, the correlation between the longitude and latitude of human viewing directions in panoramic video is rather small. This completes the analysis of *Finding 1*.

Finding 2: When watching panoramic video, subjects view the front region near the equator much more frequently than other regions.

Figure 1 shows the heat maps of viewing directions for some panoramic video sequences, as obtained from all 40 subjects. Note that the heat maps in Figure 1 have been converted from spherical coordinates to a plane for panoramic video sequences [29]. We can see from this figure that most viewing directions fall into small regions located in the front region near the equator. Furthermore, we calculate the viewing directions belonging to different regions of the panoramic videos. Since *Finding 1* illustrated that the longitude and latitude of viewing directions are almost uncorrelated with each other, it is reasonable to separately model the distribution of viewing directions along with longitude and latitude. To this end, Figure 2 shows the scatter diagrams of viewing direction frequency along with longitude and latitude, averaged over all subjects and all panoramic video sequences. In this figure, Gaussian mixture fitting curves are also plotted. According to this figure, we can see that subjects tend to watch regions near the front and equator regions, far more often than the back and pole regions. This completes the analysis of *Finding 2*, which is similar to the conclusion of [12].

Finding 3: In general, there exists high consistency in the viewed regions across different subjects for panoramic video.

We randomly and equally divide all 40 subjects into two non-overlapping groups, *A* and *B*. Then, we generate heat maps of viewing directions at one panoramic video frame for Groups *A* and *B*, which are denoted as \mathbf{H}_A and \mathbf{H}_B , respectively. Note that the heat maps for \mathbf{H}_A and \mathbf{H}_B are in plane coordinates, in which the panoramic video has been projected from sphere to plane. Here, we quantify the correlations of the heat maps of \mathbf{H}_A and \mathbf{H}_B using a linear correlation coefficient (CC) [30]. Specifically, CC is calculated by

$$\text{CC}(\mathbf{H}_A, \mathbf{H}_B) = \frac{\sum_{x,y} (\mathbf{H}_A(x,y) - \mu(\mathbf{H}_A)) \cdot (\mathbf{H}_B(x,y) - \mu(\mathbf{H}_B))}{\sqrt{\sigma(\mathbf{H}_A)^2 \cdot \sigma(\mathbf{H}_B)^2}}, \quad (3)$$

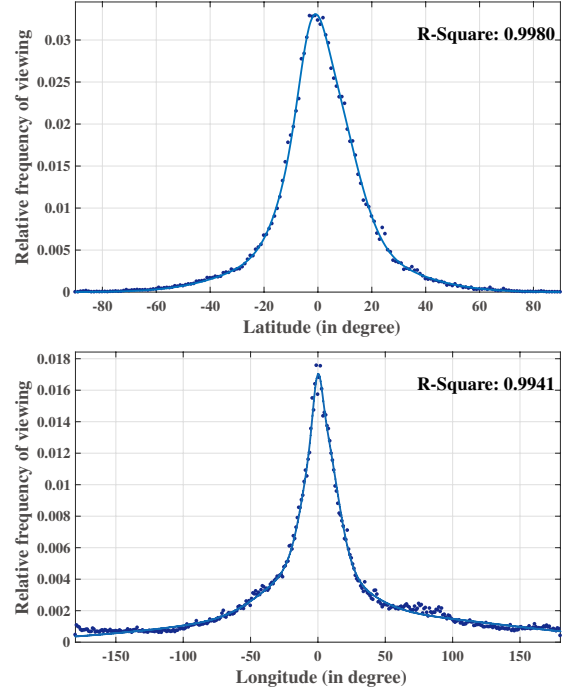


Fig. 2. Viewing direction frequency along with longitude and latitude.

where (x, y) is the pixel coordinates, and $\mu(\cdot)$ and $\sigma(\cdot)$ are the mean and standard deviation of the corresponding heat maps, respectively. A CC (ranging in $[-1, 1]$) close to +1 indicates a high consistency between the heat maps \mathbf{H}_A and \mathbf{H}_B . Table II reports the mean values and standard deviations of the CC of the viewing direction heat maps for each sequence, between Groups *A* and *B*. We can see from this table that the CC values are sufficiently high across different sequences. We can also see from this table that the CC value averaged over all 48 panoramic video sequences is 0.745, with a standard deviation of 0.114. Thus, it is clear that the subjects behaved consistently when watching panoramic video. This completes the analysis of *Finding 3*.

Finding 4: The viewing directions of different subjects are consistent in different regions according to content of panoramic video, despite being more likely to be attracted by equator and front regions.



Fig. 3. Viewing direction heat maps for selected frames from a few panoramic video sequences, in which subjects are attracted by other regions.

The scatter diagrams of Figure 2 also reveal that the regions other than the front and equator, still have potential in attracting human attention. Figure 3 demonstrates that the selected frames of several panoramic video sequences and their corresponding heat maps of viewing directions. We can see from Figure 3 that the viewing directions may focus on different regions of panoramic video rather than the front equator, depending on the video content. For example, Figure 3(c) shows that viewing directions concentrate on the corridor and people at the left hand side. This completes the analysis of this finding.

IV. SUBJECTIVE VQA METHOD

In this section, we introduce our subjective VQA method for panoramic video. In Section IV-A, we present the general configuration of the subjective test for our VQA method. In Section IV-B, the procedure of the subjective test is discussed for rating the raw quality scores of each panoramic video sequence. In Section IV-C, O-DMOS and V-DMOS are proposed as the metrics to assess subjective quality of panoramic video, which are based on the raw scores of reference and impaired panoramic videos.

A. Test configuration

Panoramic video differs from 2D video in the playing devices, the viewing experience of subjects, etc. Thus, we design the test configuration for the subjective test on assessing panoramic video, which differs from the test for 2D video. In the following, we present the general configuration of the subjective test, including display devices and the setup for subjects.

Display devices. An HMD with a corresponding video player is used to display panoramic video, rather than flat screens for displaying 2D video. This configuration is because most panoramic videos are viewed by wearing an HMD. In this paper, we use the HTC Vive as the display device of HMD and the software VD as the panoramic video player. Additionally, VD is also used to project the graphical user interface (GUI) of our quality rating software, allowing the subjects to rate panoramic video without taking off the HMD. Since panoramic video can be viewed from different viewing directions, a swivel chair is provided to subjects when viewing the panoramic videos.

Subjects. According to *Finding 3*, the viewing directions of subjects are highly consistent. Therefore, fixing the viewing regions of panoramic video in [6] is not necessary. Instead, subjects are able to freely view all content of panoramic video in our subjective test. This way, our method satisfies daily

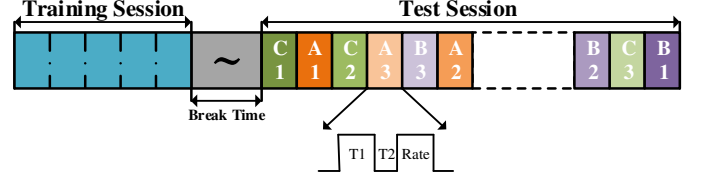


Fig. 4. Structure of the test procedure with two sessions. The SSCQS procedure is illustrated in part of the test session. A_i , B_i and C_i represent various original and impaired sequences from different contents A, B and C, respectively.

visual experience that subjects are free to access all parts of panoramic video. In addition, the initialization of viewing direction is required when watching panoramic video, which is different from viewing 2D video. In our test, the viewing directions of all subjects are initialized to be the center of front region in panoramic video, as *Finding 2* finds that subjects are more likely to be attracted by this region. However, there still exists slight inconsistency of viewed regions in panoramic video as analyzed in *Findings 3* and 4. Thus, more subjects should be involved in the subjective test for rating quality of panoramic video, than at least 15 subjects required in [9]. We recommend that at least 20 subjects are required for rating quality scores of panoramic video, as verified in Section VI-A.

B. Test procedure

Training and test. Generally speaking, the test procedure of our subjective VQA method comprises two sessions, the training and test sessions, as shown in Figure 4. The training session is introduced, as some subjects may be unfamiliar with viewing panoramic video. In the training session, subjects are told about the goal of our test. Then, they watch a group of training sequences at different quality in order to become familiar with panoramic video and its quality. Afterwards, a short break is required before entering the test session. In the test session, each sequence is displayed followed by a 3-second mid-grey screen. Compared with viewing 2D video, subjects are more likely to incur eye fatigue and motion sickness when watching panoramic video. Thus, the maximum duration of test session is 30 minutes [9]. If the test sequences last more than 30 minutes, a short break (at least 3 minutes) with the HMD taken off is added in the test session.

Quality rating. In the subjective test, SSCQS is adopted as shown in Figure 4, which means that panoramic video sequences are displayed in a random order and that sequences with the same content at different quality need to be avoided for two successive sequences. The reason for choosing SSCQS is that the subjects may continue to view unseen regions

when viewing panoramic video with the same content, which differs from the viewing characteristics of 2D video. After viewing each sequence, subjects are required to rate its quality. The grading scores in the test session are achieved using a continuous-scale slider with a cursor in our quality rating GUI. The score Q has a range from 0 to 100, in the form of 5 levels: excellent ($80 \leq Q \leq 100$), good ($60 \leq Q < 80$), fair ($40 \leq Q < 60$), poor ($20 \leq Q < 40$) and bad ($0 \leq Q < 20$).

Data collection. There are two kinds of data to be collected and processed, including the raw subjective quality scores of the panoramic video sequences as mentioned above. The other is the viewing direction data of subjects during sequence playback, which relate the quality score to the regions of panoramic video that were viewed. This also enables the calculation of V-DMOS, to be discussed next.

C. Processing of subjective scores

O-DMOS. Given the raw quality scores for each sequence, we follow the DMOS calculation method of 2D video as detailed in [17] to compute the O-DMOS, which indicates the overall quality of each panoramic video sequence. Specifically, the difference in the quality scores between the reference and impaired sequences is calculated for each subject. Let S_{ij} and S_{ij}^{ref} denote the raw subjective scores assigned by subject i to sequence j and the corresponding reference sequence, respectively. Then, the difference score d_{ij} can be simply obtained by

$$d_{ij} = S_{ij}^{\text{ref}} - S_{ij}. \quad (4)$$

Afterwards, the difference score d_{ij} needs to be converted to a Z-score Z_{ij} [31] using

$$\mu_i = \frac{1}{M_i} \sum_{j=1}^{M_i} d_{ij}, \quad \sigma_i = \sqrt{\frac{1}{M_i - 1} \sum_{j=1}^{M_i} (d_{ij} - \mu_i)^2}, \quad (5)$$

$$Z_{ij} = \frac{d_{ij} - \mu_i}{\sigma_i}, \quad (6)$$

where M_i is the number of test sequences viewed by subject i .

Here, we need to ensure that each subject is valid by examining the Z-scores assigned by this subject. In other words, the Z-scores from invalid subjects should not be included when calculating the O-DMOS for measuring the subjective quality of panoramic video. We apply the subject rejection method [9] to remove the Z-scores of some subjects if 5% of the Z-scores assigned by these subjects fall outside the range of two standard deviations from the mean Z-scores.

Then, the Z-score Z_{ij} needs to be linearly rescaled to fall within the range of $[0, 100]$. Assume that the Z-scores of a subject follow Gaussian distribution. Then, Z_{ij} of (6) is distributed as a standard Gaussian, i.e. $\mathcal{N}(0, 1)$, in which the mean is 0 and the standard deviation is 1. Thus, 99% of the Z-scores lie in the range of $[-3, 3]$. To make such Z-scores $\in [0, 1]$, we normalize them by $(Z_{ij} + 3)/6$. Then, the normalized Z-scores are rescaled to be Z'_{ij} as follows,

$$Z'_{ij} = \frac{100(Z_{ij} + 3)}{6}, \quad (7)$$

such that 99% of the values of Z'_{ij} fall into the range of $[0, 100]$.

Finally, the O-DMOS value of sequence j is computed by averaging Z'_{ij} from N_j valid subjects:

$$\text{O-DMOS}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} Z'_{ij}. \quad (8)$$

V-DMOS. According to *Finding 3*, there is still a slight inconsistency in the panoramic video viewing directions. *Finding 4* further shows that all the regions of panoramic video can attract human attention. Thus, V-DMOS is used in our subjective VQA method to quantify the subjective quality of different regions of panoramic video, by making use of the collected raw quality scores and viewing direction data. First, we need to compute the ratio of the frequency, with which subject i views region r in sequence j , denoted as f_{ij}^r . Note that f_{ij}^r needs to be normalized to satisfy

$$\sum_r f_{ij}^r = 1. \quad (9)$$

When $f_{ij}^r > f_0$, where f_0 is a threshold, subject i (after subject rejection [9]) is added to collection \mathbf{I}_{jr} . Assuming that the size of \mathbf{I}_{jr} is $N_{\mathbf{I}_{jr}}$, the DMOS value for region r in sequence j can be obtained by

$$\text{DMOS}_{jr} = \frac{1}{N_{\mathbf{I}_{jr}}} \sum_{i \in \mathbf{I}_{jr}} Z'_{ij}. \quad (10)$$

If $\mathbf{I}_{jr} = \emptyset$, then DMOS_{jr} is an invalid value, denoted by “—”. Finally, the vector of V-DMOS can be represented by

$$[\text{O-DMOS}_j \quad \text{DMOS}_{j1} \quad \cdots \quad \text{DMOS}_{jr} \quad \cdots \quad \text{DMOS}_{jR}], \quad (11)$$

where R is the total number of regions in panoramic video. The same as [32], there are 6 regions of panoramic video in our method: front, left, back, right, top, and bottom. As a result, our V-DMOS is able to qualify both the overall and regional quality degradation for impaired panoramic video.

V. OBJECTIVE VQA METHODS

In this section, we propose two objective VQA methods for panoramic video, which are on the basis of the traditional PSNR mechanism and our findings in Section III-B. Both of these methods impose weights on the pixel-wise distortion in calculating the PSNR, according to the possibility of attracting human attention. Thus, these methods are called perceptual VQA (P-VQA) methods. The first method mainly focuses on weighting the distortion of pixels according to their locations in panoramic video rather than their contents. Thus, this method is called the non-content-based P-VQA (NCP-VQA) method, to be discussed in Section V-A. The second method assigns weights to pixel-wise distortion based on the viewing directions predicted with respect to the content of panoramic video and is thus called the content-based P-VQA (CP-VQA) method. This is to be introduced in Section V-B.

TABLE III
VALUES OF THE PARAMETERS IN (12).

k/k'	a_k	b_k	c_k	$a'_{k'}$	$b'_{k'}$	$c'_{k'}$
1	0.0034	-0.1549	4.6740	0.0075	-2.3738	6.6437
2	0.0106	1.5140	18.51	0.0209	1.8260	14.8171
3	0.0032	6.3670	110.5	0.0057	1.4618	36.1311

A. Non-content-based perceptual VQA method

According to *Finding 2*, front regions near the equator are viewed more frequently than other regions in panoramic video. Thus, it is necessary to consider such viewing direction frequency when calculating the non-content-based perceptual PSNR (NCP-PSNR) for our NCP-VQA method for panoramic video. Let $\varphi \in [-180^\circ, 180^\circ]$ and $\theta \in [-90^\circ, 90^\circ]$ denote the longitude and latitude of a viewing direction in degrees, respectively. Since *Finding 1* points out that the longitude and latitude of the viewing directions are almost independent of each other, the distribution of viewing direction frequency $u(\varphi, \theta)$ can be modeled using the following Gaussian mixture model (GMM):

$$u(\varphi, \theta) = \left\{ \sum_{k=1}^3 a_k \exp \left[-\left(\frac{\varphi - b_k}{c_k} \right)^2 \right] \right\} \left\{ \sum_{k'=1}^3 a'_{k'} \exp \left[-\left(\frac{\theta - b'_{k'}}{c'_{k'}} \right)^2 \right] \right\}. \quad (12)$$

In (12), a_k , b_k and c_k are parameters of the GMM for the viewing direction distribution in longitude, and $a'_{k'}$, $b'_{k'}$ and $c'_{k'}$ are GMM parameters for the viewing direction distribution in latitude. The values of these parameters can be obtained via least squares fitting for all viewing directions in our database, and they are reported in Table III. Note that the number of Gaussian components for the fitting of (12) is set to 3, for making the fitting error convergent. The R-square value for the fitting is 0.84, which implies that (12) is effective in modeling the frequency distribution of viewing direction in panoramic video.

Next, we take into account the equirectangular projection for our NCP-PSNR metric. Given a panoramic video under an equirectangular projection with a resolution of $W \times H$, the probability of each pixel being in the viewing direction during one frame can be obtained by [29]:

$$v(x, y) = u \left(-360 \left(\frac{x-1}{W-1} - \frac{1}{2} \right), -180 \left(\frac{y-1}{H-1} - \frac{1}{2} \right) \right), \quad (13)$$

where (x, y) is the pixel coordinates, with $1 \leq x \leq W$ and $1 \leq y \leq H$. Note that our VQA method can be easily extended to other projections by replacing the projection formulation.

In fact, pixels within a viewport centered in one viewing direction are all accessible to subjects. Thus, the pixels within a viewport should be of equal importance in evaluating the quality of panoramic video. To model possible viewports, we need to generate the non-content-based weight map in our NCP-VQA method, based on the probability of viewing direction, i.e., $v(x, y)$ in (13). Assume that $\mathbf{P}_{x,y}$ is the viewport, the center of which is the viewing direction (x, y) .

According to studies on near peripheral vision [33], the ranges of $\mathbf{P}_{x,y}$ are set to $[-30^\circ, 30^\circ]$ in both directions. For each pixel (s, t) in a panoramic video frame, we can find all viewports including this pixel, and the viewing directions of these viewports constitute a collection denoted by $\mathbf{V}_{s,t}$. Then, the non-content-based weight map can be obtained by

$$w(s, t) = \max_{(s', t') \in \mathbf{V}_{s,t}} v(s', t'). \quad (14)$$

Afterwards, the non-content-based weight map needs to be normalized by

$$\tilde{w}(s, t) = \frac{w(s, t)}{\sum_{s,t} w(s, t)}, \quad (15)$$

to satisfy

$$\sum_{s,t} \tilde{w}(s, t) = 1. \quad (16)$$

Finally, based on the definition of the PSNR, the NCP-PSNR for each panoramic video frame can be calculated¹ as

$$\text{NCP-PSNR} = 10 \log \frac{I_{\max}^2}{\sum_{s,t} (I(s, t) - I'(s, t))^2 \cdot \tilde{w}(s, t)}, \quad (17)$$

where $I(s, t)$ and $I'(s, t)$ are intensities of pixel (s, t) in the reference and processed panoramic videos, respectively. Additionally, I_{\max} is the maximum intensity value of the videos (=255 for 8-bit intensity).

B. Content-based PVQA method

Finding 4 shows that the viewing directions of the subjects are also correlated with the contents of the panoramic video. According to this finding, we further develop a CP-VQA method for panoramic video, in which the content-based perceptual PSNR (CP-PSNR) is measured. Figure 5(a) summarizes the procedure of our CP-VQA method. As shown in this figure, our CP-VQA method consists of two parts: the model training and CP-PSNR calculation. For the model training, the input includes panoramic video frames and their corresponding viewing directions for all subjects. Then, a random forest model of classification is trained to predict viewing directions. For the CP-PSNR calculation, each panoramic video frame is taken as the input. After extracting several candidates from the input frame, a viewing direction can be predicted using maximum a posteriori (MAP) estimation for the incoming frames, with regard to detected features and the trained model. Given the predicted viewing direction, the viewport binary map is calculated. Then, the content-based weight map is generated by multiplying the viewport binary map with the non-content-based weight map. Finally, the CP-PSNR is obtained by imposing the content-based weight map in the PSNR. In the following, we present the details of our CP-VQA method.

In our CP-VQA method, the first step is to extract viewing direction candidates for the next panoramic video frame. Given the input panoramic video frame and its current viewing direction, the procedure for extracting the viewing direction

¹Because of (15) and (16), we do not need to divide NCP-PSNR in (17) by the number of pixels.

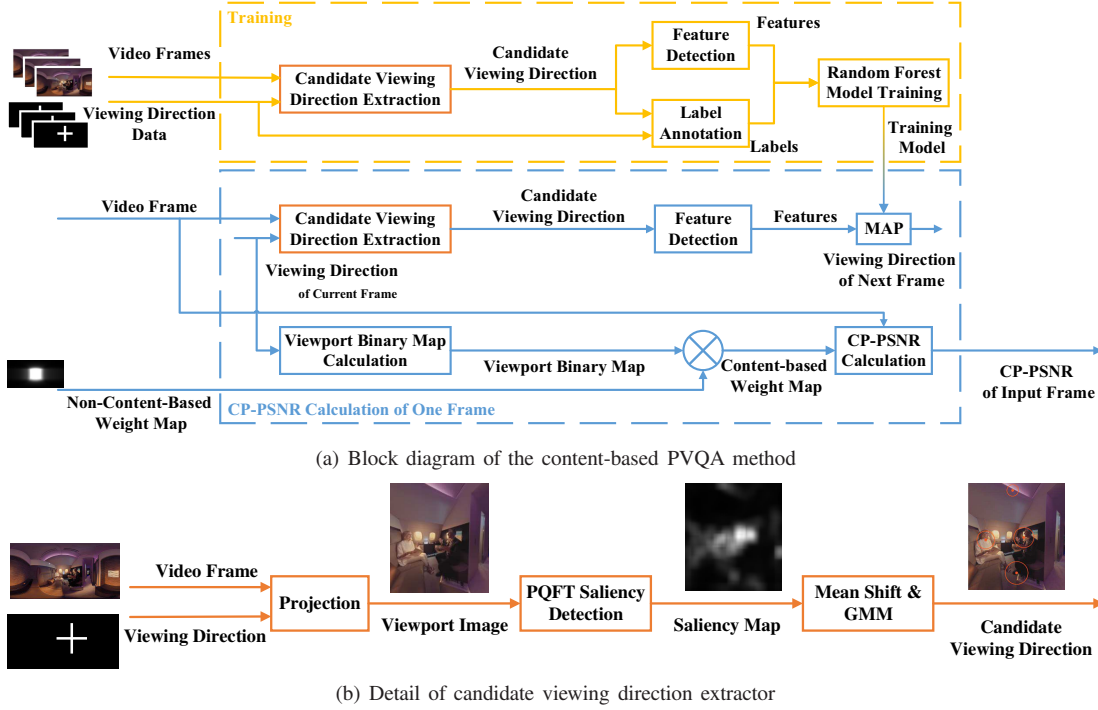


Fig. 5. Viewing direction heap maps for selected frames of a few panoramic video sequences, in which subjects are attracted by other regions.

candidates is shown in Figure 5(b). To extract the viewing direction candidates, a viewport projection [12] is applied to obtain the viewport image of the input video frame given the current viewing direction. Note that the viewport image is regarded as the image that subjects can actually see through the HMD. Subsequently, a saliency map of the viewport image is generated using a simple yet effective saliency detection algorithm, phase spectrum of quaternion Fourier transform (PQFT) [34]. Then, the mean-shift [35] and Gaussian mixture model (GMM) are used to select the salient regions with the potential for attracting human attention, as candidates of the viewing directions for the incoming frames.

Regarding the model training, the inputs are video frames and their corresponding viewing directions for each subject, which are derived from our viewing direction database for panoramic video. Then, several candidates of viewing directions are obtained using the aforementioned extractor. For each candidate, the features, which are correlated with the viewing direction transition from the current one to the candidate, are detected to train a random forest classifier [36]. The features include (1) the Euclidean distance from the viewport center to the candidate; (2) the angle between the viewport center and the candidate; (3) the standard deviation of the GMM used for extracting the candidate; (4) the averaged saliency value of the region around the candidate; and (5) the local intensity contrast of the neighborhood around the candidate [37]. These features form a vector \mathbf{v} , which is the input to the random forest classifier. Furthermore, the viewing direction for the same subject in the next frame is annotated as the ground-truth and is the target output of the random forest classifier. Finally, the random forest classifier can be trained with the feature vectors and ground-truth viewing directions, and is

then used to calculate the CP-PSNR of each panoramic video frame.

For the CP-PSNR calculation, the first step is to predict the viewing direction. To predict the viewing direction, a few viewing direction candidates are extracted. Then, the MAP estimation is employed to select one viewing direction from the candidates given the detected features embedded in vector \mathbf{v} and the trained random forest model. Specifically, assuming that C is a viewing direction candidate, the averaged posterior probability of candidate C being the viewing direction (i.e., belonging to the positive class) can be obtained by

$$g_{\lambda_+}(C) = \frac{1}{T} \sum_{\tau=1}^T P(\lambda_+ | \mathbf{v}_\tau(C)), \quad (18)$$

where λ_+ represents the positive class, and $\mathbf{v}_\tau(C)$ is the feature vector of C input to tree τ . In addition, T is the number of classification trees in the trained random forest model. Note that each tree randomly chooses some features from the input feature vector \mathbf{v} for the classification, such that $\mathbf{v}_\tau(C) \subseteq \mathbf{v}$. Finally, the viewing direction is predicted for the next frame by MAP as follows,

$$V = \underset{C}{\operatorname{argmax}} g_{\lambda_+}(C). \quad (19)$$

Given the predicted viewing direction, a viewport binary map can be generated, in which 1 indicates that the corresponding pixel is in the viewport range and 0 means that the pixel is out of the viewport range. Then, a content-based weight map $w'(s, t)$ is generated via multiplying the viewport binary map by the non-content-base weight map

TABLE IV
THE FINAL OUTPUT AS THE V-DMOS OF THE IMPAIRED TEST SEQUENCES.

QP	Name	V-DMOS	Name	V-DMOS	Name	V-DMOS	Name	V-DMOS
27	Dianying	[43,43,45,36,43,48,33]	Fengjing1	[36,37,37,36,34,—,—]	Fengjing3	[36,35,34,43,38,72,21]	Hangpai1	[33,33,30,28,29,—,34]
37		[65,65,64,69,66,—,57]		[64,64,65,70,66,64,—]		[43,44,47,41,47,—,35]		[47,47,48,41,46,—,41]
42		[71,71,66,64,71,—,54]		[70,70,70,60,70,—,55]		[54,55,54,44,54,—,38]		[58,58,53,65,52,—,51]
27	Hangpai2	[33,33,32,34,34,—,19]	Hangpai3	[36,36,35,33,36,—,—]	Tiyu1	[43,43,40,42,39,—,—]	Tiyu2	[35,35,31,—,30,—,—]
37		[40,40,40,43,40,—,32]		[47,47,44,48,48,—,46]		[59,59,56,—,60,—,66]		[58,57,59,60,61,—,70]
42		[52,52,54,37,51,—,48]		[58,57,55,63,62,—,—]		[70,71,66,67,65,55,—]		[66,66,64,—,66,—,71]
27	Xinwen1	[34,33,33,34,33,—,35]	Xinwen2	[34,34,33,34,34,—,46]	Yanchanghui1	[35,34,35,42,34,—,—]	Yanchanghui2	[34,33,33,32,34,—,—]
37		[47,46,48,47,47,—,—]		[55,55,55,51,56,—,59]		[45,43,46,59,43,43,—]		[52,50,52,58,55,—,—]
42		[58,57,61,72,57,—,50]		[67,67,66,59,67,—,70]		[59,58,62,65,58,—,—]		[62,62,62,58,64,63,—]

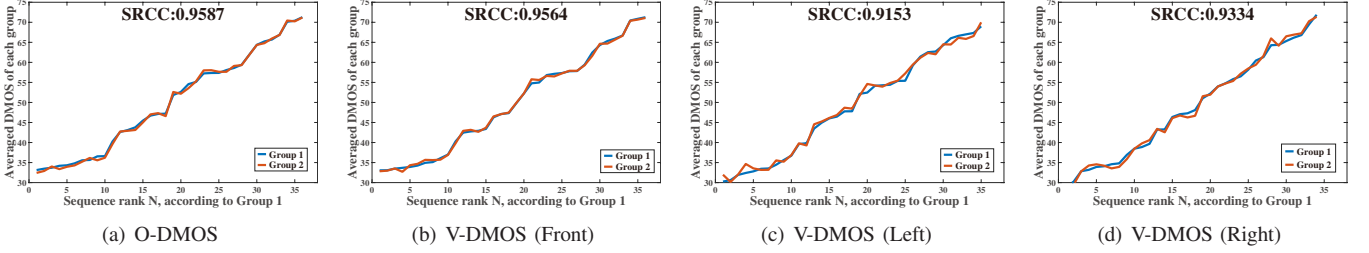


Fig. 6. Curves of the O-DMOS/V-DMOS values of each impaired sequence for two non-overlapping groups of subjects with equal size, in which the sequences are ranked in increasing order according to the O-DMOS/V-DMOS values of Group 1.

$\tilde{w}(s, t)$ (introduced in Section V-A). We further normalize $w'(s, t)$ by

$$\tilde{w}(s, t) = \frac{w'(s, t)}{\sum_{s, t} w'(s, t)}. \quad (20)$$

Finally, the CP-PSNR of each panoramic video frame can be calculated as

$$\text{CP-PSNR} = 10 \log \frac{I_{\max}^2}{\sum_{s, t} (I(s, t) - I'(s, t))^2 \cdot \tilde{w}(s, t)}. \quad (21)$$

As a result, CP-VQA can be obtained for measuring the objective quality of panoramic video.

VI. EXPERIMENTAL RESULTS

A. Validation on our subjective VQA method

Test benchmark and setting. In this section, we validate the effectiveness of our subjective VQA method. First, all 12 uncompressed panoramic video sequences from [38] (in YUV 4:2:0 format at resolution 4096×2048) are selected as the references. The duration of these sequences is all 12 seconds with a frame rate of 25 frame per second (fps). Then, H.265/HEVC is used to compress these 12 sequences at 3 different bit-rates, under an equirectangular projection. For each sequence, the 3 bit-rates are set to be the actual bit-rates by QP = 27, 32 and 37. Thus, there are 12 reference and 36 impaired sequences for the test in total. Note that all test sequences are non-overlapping with 48 sequences of our viewing direction database introduced in Section III-A.

A total of 48 subjects participated in the subjective test for our VQA method (presented in Section IV). In the test, subjects were required to view and rate all sequences for raw subjective scores. Next, the O-DMOS and V-DMOS are calculated with the rated raw scores. Here, we simply set the threshold f_0 to be $1/6$ in the V-DMOS calculation, as there

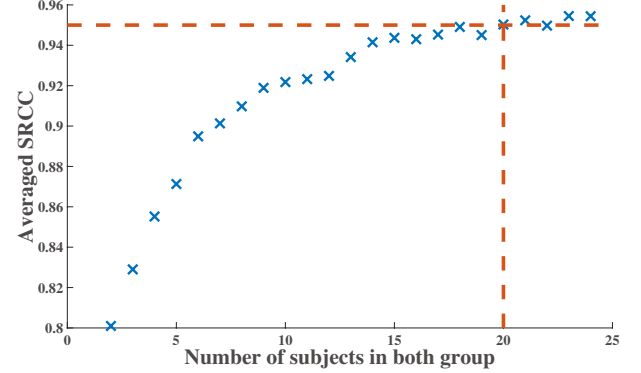


Fig. 7. SRCC of O-DMOS scores between two groups with increasing numbers of subjects in both groups, representing averaged result over 30 trials.

are 6 regions in our panoramic videos. It is worth mentioning that no subject was rejected in the calculation of the O-DMOS and V-DMOS values after using the subject rejection scheme from [9]. Finally, the values of the O-DMOS and V-DMOS obtained from the raw quality scores of the 48 subjects are reported in Table IV².

Evaluation on the effectiveness of our subjective VQA method. The effectiveness of our subjective VQA method is verified by evaluating the correlations between the O-DMOS/V-DMOS scores of different groups of subjects. Specifically, all 48 subjects are randomly and equally divided into two non-overlapping groups, Group 1 and Group 2, by 30 trials. Then, the O-DMOS/V-DMOS values are averaged over 30 trials, and the correlations of the averaged O-DMOS/V-DMOS values between two groups are evaluated as follows.

²Note that the O-DMOS is included in the V-DMOS as the first element and in bold in Table IV. The second to the seventh elements represent the DMOS scores of the front, left, back, right, top, and bottom regions, respectively.

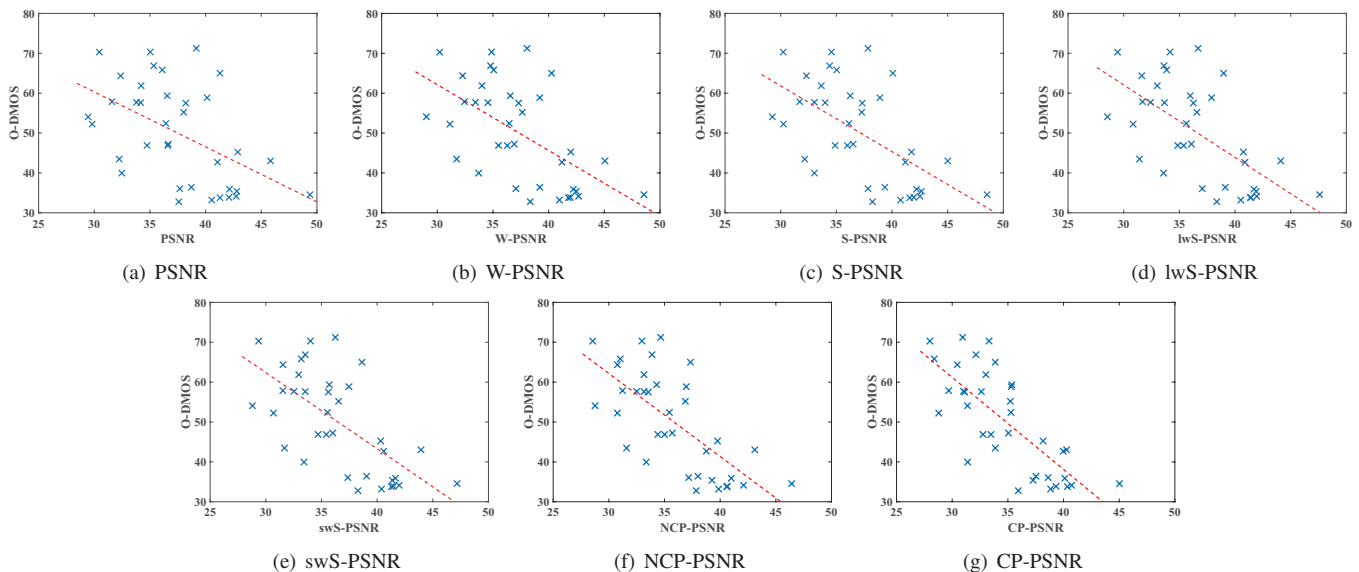


Fig. 8. Scatter plots of the objective VQA results versus the O-DMOS values for all 36 impaired sequences.

TABLE V
SRCC BETWEEN THE O-DMOS AND V-DMOS SCORES OF DIFFERENT REGIONS.

Region	Front	Left	Right	Back	Top	Bottom
SRCC	0.9972	0.9794	0.9750	0.8844	-0.0857	0.8487

Figure 6 shows the curves of the ranked O-DMOS/V-DMOS values³ for all 36 impaired sequences obtained by Group 1, and the figure also presents the O-DMOS/V-DMOS values by Group 2 for the sequences ranked by the values for Group 1. We can see from this figure that the correlations between two groups of O-DMOS/V-DMOS values are extremely high. We quantify such correlations using Spearman’s rank correlation coefficient (SRCC), which is shown in Figure 6. The high SRCC values again indicate the agreement between the two groups for O-DMOS and V-DMOS. Since two randomly selected groups can be seen as the results from two subjective tests, the achieved agreement over different subjective tests implies that our method is effective in assessing subjective quality of panoramic video.

Performance analysis of our subjective VQA method.

It is necessary to ascertain the minimum number of subjects required for our subjective VQA method. To this end, we measure the SRCC of the O-DMOS values between the two groups with different numbers of subjects. Accordingly, Figure 7 shows the SRCC with an increased number of subjects in Groups 1 and 2, which is the averaged result over 30 trials. We can see that SRCC converges when the number of subjects is more than 20. Thus, we recommend 20 as the minimum number of subjects for our VQA method.

It is also interesting to investigate the relationship between the O-DMOS and V-DMOS values of different regions. Table V shows the SRCC results between the O-DMOS and V-DMOS values of different regions, which are calculated

from all 48 subjects. It is clear that the V-DMOS values of the front, left and right regions have strong correlation with the O-DMOS values. In contrast, the V-DMOS values of the back and bottom regions are generally correlated with the O-DMOS values. However, the SRCC result for the V-DMOS values of the top region is rather small. It is because the V-DMOS values of the top region are determined by only few subjects, since most subjects pay no attention to the top region. In general, there exists a high correlation between O-DMOS and V-DMOS, verifying the effectiveness of the proposed V-DMOS metric.

B. Validation on our objective VQA methods

Test benchmark and evaluation metrics. The performance of our objective VQA methods is evaluated by measuring the agreement between subjective and objective quality. The performance evaluation is conducted on 36 impaired sequences of 12 uncompressed panoramic video sequences, as mentioned in Section VI-A. Here, the subjective quality of those impaired sequences is the O-DMOS values of 48 subjects obtained in Section VI-A. For calculating CP-PSNR, all 48 panoramic video sequences from our viewing direction database presented in Section III-A, which does not overlap with any test sequence of this section, are used as the training data to learn the random forest model. All PSNR-related objective metrics are calculated on the Y component and averaged over all frames for each impaired sequence. Given the O-DMOS results, the performance of the objective VQA is measured with SRCC, Pearson correlation coefficient (PCC), Root-Mean-Square Error (RMSE) and Mean Absolute Error

³Due to space limitations, we only show the values of the front, left and right regions for V-DMOS.

TABLE VI
COMPARISON OF THE PERFORMANCES OF OBJECTIVE VQA METHODS

metrics	PSNR	W-PSNR [6]	S-PSNR [12]	lwS-PSNR [12]	swS-PSNR [12]	NCP-PSNR (our)	CP-PSNR (our)
SRCC	0.5117	0.5748	0.5897	0.6180	0.6366	0.7019	0.7506
PCC	0.5075	0.5795	0.5902	0.6359	0.6582	0.6961	0.7559
RMSE	88.153	87.983	87.795	87.232	87.074	86.421	85.399
MAE	87.488	87.357	87.177	86.635	86.486	85.845	84.848

(MAE). SRCC measures the monotonicity of the objective quality with respect to subjective quality, while PCC quantifies the correlation coefficients between subjective and objective quality. In addition, RMSE and MAE measure the difference between the objective and subjective VQA results. Obviously, large-valued SRCC and PCC, or a small-valued RMSE and MAE, indicate a high correlation between objective and subjective metrics.

Comparison of scatter plots. Now, we compare the two objective VQA metrics of our method (NCP-PSNR and CP-PSNR) with traditional PSNR and four state-of-the-art metrics. The four metrics include S-PSNR, latitude-weighted S-PSNR (lwS-PSNR) and sphere-weighted S-PSNR (swS-PSNR), all of which are from [12], as well as the latest W-PSNR proposed in [6]. Figure 8 shows the scatter plots of objective VQA results versus the O-DMOS results for all 36 impaired sequences along with the minimal linear fitting curves. In general, intensive scatter points close to the fitting curve indicate an high correlation of the objective VQA results with the subjective results, validating the effectiveness of the objective VQA method. It can be clearly seen from Figure 8 that the VQA results of our NCP-PSNR and CP-PSNR metrics have a much higher correlation with the O-DMOS results, compared to other VQA methods. Therefore, we can conclude that both the NCP-PSNR and CP-PSNR perform far better than other metrics.

Comparison on quantification results. Furthermore, Table VI reports the SRCC, PCC, RMSE and MAE between the results of five objective VQA metrics and the subjective O-DMOS results, over all 36 impaired panoramic video sequences. Here, the absolute values of SRCC and PCC are reported, since the correlation between the objective VQA results and the O-DMOS results is negative. As can be seen in Table VI, the NCP-PSNR and CP-PSNR of our VQA methods significantly outperform other methods in terms of SRCC, PCC, RMSE and MAE. In addition, we can see from Table VI that CP-PSNR is superior to the NCP-PSNR, with an increase in SRCC/PCC and a reduction in RMSE and MAE. This implies that the viewing directions predicted with regard to the video content are effective in improving the performance of objective VQA on panoramic video.

VII. CONCLUSION

In this paper, we have proposed both subjective and objective VQA methods for evaluating the quality degradation of impaired panoramic video. In contrast with the conventional VQA methods, human viewing directions were investigated and then taken into account in our VQA methods. Specifically,

we conducted an experiment to establish a new database, which contains the viewing directions from 40 subjects on viewing 48 panoramic video sequences. Next, we found from our database that subjects consistently prefer to looking at the center of front region of panoramic video, but there still exists dependency on video content for viewing directions. In light of our findings, we proposed two subjective VQA metrics, O-DMOS and V-DMOS, measuring the overall and regional quality reduction of impaired panoramic video, respectively. In addition, we proposed two objective metrics, NCP-PSNR and CP-PSNR, for assessing the quality loss of impaired panoramic videos. In NCP-PSNR, the quality loss is weighed according to statistical results on the preference for the center of the front region, while CP-PSNR imposes quality loss using weights with respect to possible viewing directions predicted upon the video content. Finally, our experimental results validate the effectiveness of our subjective and objective VQA methods.

There are two promising directions for future work. First, our objective VQA methods only work on PSNR-rated metrics for panoramic video. Other advanced metrics, such as structural similarity index (SSIM), can also incorporate possible viewing directions in measuring the quality of panoramic video. Second, future work may apply our VQA method in optimizing the encoder of panoramic video. For example, NCP-PSNR or CP-PSNR can be maximized in the bit allocation when encoding panoramic video.

REFERENCES

- [1] HUAWEI iLab, "VR data report," HUAWEI Report, 2016, <https://mp.weixin.qq.com/s/tcsm9NIECa7d1L7gZekrrQ>.
- [2] W. Sarmiento and C. Quintero, "Panoramic immersive videos-3d production and visualization framework," in *International Conference on Signal Processing and Multimedia Applications*, 2009, pp. 173–177.
- [3] R. Konrad, E. A. Cooper, and G. Wetzstein, "Novel optical configurations for virtual reality: evaluating user preference and performance with focus-tunable and monovision near-eye displays," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 1211–1220.
- [4] the MPEG Virtual Reality Ad-hoc Group, "Summary of survey on virtual reality," in *ISO/IEC JTC 1/SC 29/WG 11 N16542*, 2016.
- [5] E. Upenik, M. Řeřábek, and T. Ebrahimi, "Testbed for subjective evaluation of omnidirectional visual content," in *Picture Coding Symposium*. IEEE, 2016, pp. 1–5.
- [6] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," in *SPIE Optical Engineering+ Applications*, 2016, pp. 99700C–99700C.
- [7] Recommendation, ITU-R, "BT. 710-4: Subjective assessment methods for image quality in high-definition television," Technical Report, ITU-R, Tech. Rep., 1998.
- [8] Recommendation, ITUT, "P. 910: Subjective video quality assessment methods for multimedia applications," *ITU, Geneva*, 2008.
- [9] Assembly, ITU-R, "Methodology for the subjective assessment of the quality of television pictures," 2012.

- [10] T. K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J.-R. Ohm, and G. J. Sullivan, "Video quality evaluation methodology and verification testing of hevcc compression performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 76–90, 2016.
- [11] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "A subjective study to evaluate video quality assessment algorithms," in *IS&T/SPIE Electronic Imaging*, 2010, pp. 75 270H–75 270H.
- [12] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 2015, pp. 31–36.
- [13] M. Xu, C. Li, Y. Liu, X. Deng, and J. Lu, "A subjective visual quality method of panoramic videos," in *2017 IEEE International Conference on Multimedia and Expo*. IEEE, 2017.
- [14] C. Li, M. Xu, Y. Liu, M. Bai, and W. Wei, "IEEE1857.9-N1013 A subjective evaluation methodology on panoramic video," IEEE1857.9 7th Meeting: Hainan, China, 2016.
- [15] M. H. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," in *Visual Communications and Image Processing 2003*, 2003, pp. 573–582.
- [16] C. Lee, H. Choi, E. Lee, S. Lee, and J. Choe, "Comparison of various subjective video quality assessment methods," in *Electronic Imaging 2006*, 2006, pp. 605 906–605 906.
- [17] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, 2010.
- [18] P. Pourashraf, F. Safaei, and D. R. Franklin, "Minimisation of video downstream bit rate for large scale immersive video conferencing by utilising the perceptual variations of quality," in *IEEE International Conference on Multimedia and Expo*. IEEE, 2014, pp. 1–6.
- [19] Recommendation, ITU-R, "BT. 2021-1: Subjective methods for the assessment of stereoscopic 3DTV systems," *ITU-R*, vol. 2021, 2015.
- [20] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Transactions on Multimedia*, vol. 4, no. 1, pp. 129–132, 2002.
- [21] C.-H. Chou and C.-W. Chen, "A perceptually optimized 3-D subband codec for video communication over wireless channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 2, pp. 143–156, 1996.
- [22] A. Cavallaro, O. Steiger, and T. Ebrahimi, "Semantic video analysis for adaptive content delivery and automatic description," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 10, pp. 1200–1209, 2005.
- [23] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image and Vision Computing*, vol. 29, no. 1, pp. 1–14, 2011.
- [24] Z. Chen and C. Guillemot, "Perceptually-friendly H. 264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 6, pp. 806–819, 2010.
- [25] M. Xu, X. Deng, S. Li, and Z. Wang, "Region-of-interest based conversational HEVC coding with hierarchical perception model of face," *IEEE Journal of Selected Topics on Signal Processing*, vol. 8, no. 3, pp. 475–489, Jun. 2014.
- [26] N. Liu and G. Zhai, "Free energy adjusted peak signal to noise ratio (FEA-PSNR) for image quality assessment," *Sensing and Imaging*, vol. 18, no. 1, p. 11, 2017.
- [27] T. Na and M. Kim, "A novel no-reference PSNR estimation method with regard to deblocking filtering effect in H. 264/AVC bitstreams," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 2, pp. 320–330, 2014.
- [28] E. Upenik, M. Rerabek, and T. Ebrahimi, "On the performance of objective metrics for omnidirectional visual content," in *9th International Conference on Quality of Multimedia Experience*, no. EPFL-CONF-227464, 2017.
- [29] J. P. Snyder, *Map projections—A working manual*. US Government Printing Office, 1987, vol. 1395.
- [30] J. Li, C. Xia, Y. Song, S. Fang, and X. Chen, "A data-driven metric for comprehensive evaluation of saliency models," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 190–198.
- [31] A. M. Van Dijk, J.-B. Martens, and A. B. Watson, "Quality assessment of coded images using numerical category scaling," in *Advanced Networks and Services*. International Society for Optics and Photonics, 1995, pp. 90–101.
- [32] E. Praun and H. Hoppe, "Spherical parametrization and remeshing," in *ACM Transactions on Graphics*, vol. 22, no. 3. ACM, 2003, pp. 340–349.
- [33] J. Besharse and D. Bok, *The retina and its disorders*. Academic Press, 2011.
- [34] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 185–198, 2010.
- [35] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.
- [36] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [37] D. Rudoy, D. B. Goldman, E. Shechtman, and L. Zelnik-Manor, "Learning video saliency from human gaze using candidate selection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1147–1154.
- [38] IEEE1857.9 1st Meeting: Beijing, China, "1857.9-01-N0001 output document," 2016.