A SUBJECTIVE VISUAL QUALITY ASSESSMENT METHOD OF PANORAMIC VIDEOS

Mai Xu[†*], *Chen Li*[†], *Yufan Liu*[†], *Xin Deng*[‡] *and Jiaxin Lu*[†]

[†]School of Electronic and Information Engineering, Beihang University, China
[‡]Communications and Signal Processing Group, Imperial College London, London SW7 2AZ, UK.

ABSTRACT

Different from 2-dimensional (2D) videos, panoramic videos contain spherical viewing direction with the support of head-mounted displays, thus improving immersive and interactive visual experience. Unfortunately, to our best knowledge, there are few subjective visual quality assessment (VQA) methods for panoramic videos. In this paper, we therefore propose a subjective VQA method for assessing quality loss of impaired panoramic videos. Specifically, we first establish a database containing viewing direction data of several subjects on watching panoramic videos. Then, we find out that there exists high consistency of viewing direction on panoramic videos across different subjects. Upon this finding, we present a procedure of subjective test in measuring quality of panoramic videos by different subjects, yielding different mean opinion score (DMOS). To couple with inconsistency of viewing directions on panoramic videos, we further propose a vectorized DMOS metric. Finally, experimental results verify that our subjective VQA method, in the forms of both overall and vectorized DMOS metrics, is effective in measuring subjective quality of panoramic videos.

***Index Terms*—** Panoramic videos, consistency, visual quality assessment (VQA), subjective quality

1. INTRODUCTION

Virtual reality (VR) has been rapidly developing in recent years. As an important development direction of VR, panoramic videos [1] provide a $360 \times 180^\circ$ field of view (FoV). As such, with the support of head-mounted displays (HMD), immersive and even interactive visual experience [2] can be achieved through panoramic videos, which is far advanced than traditional 2-dimensional (2D) videos. However, it is likely that the quality of experience (QoE) [3] of panoramic videos dramatically degrades due to compression artifacts or low resolutions. Such QoE degradation may make subjects feel uncomfortable, according to the survey of MPEG [4]. Therefore, there is a pressing need for visual quality assessment (VQA) on impaired panoramic videos.

Both objective and subjective methods are necessary for VQA on panoramic videos. For objective VQA, the early work of [1, 5] takes into account the spherical characteristic of panoramic videos. For example, Yu *et al.* [5] proposed a sphere-based peak signal to noise ratio (S-PSNR), which calculates PSNR on a set of uniformly sampled points on sphere, instead of rectangular-mapped pixels. By applying interpolation algorithms, S-PSNR is able to cope with objective quality assessment for panoramic videos under different projections. Besides, Zakharchenkoa *et al.* [1] proposed a weighted PSNR by using gamma-corrected pixel values for PSNR and SSIM calculation process.

Unfortunately, to our best knowledge, there are few specific subjective VQA methods for measuring quality reduction of panoramic videos. In contrast, there are a great number of subjective VQA methods for 2D videos. Over the past two decades, the international telecommunication union (ITU) has recommended several VQA methodologies for traditional 2D videos [6–8]. Among those recommendations, double stimulus continuous quality scale (DSCQS) [9], single stimulus continuous quality scale (SSCQS) [10] and single stimulus continuous quality evaluation (SSCQE) [11] were adopted to determine the displaying orders of sequences during the subjective test for VQA on 2D videos. Besides, two metrics are widely used in rating subjective VQA on 2D videos: one is the mean opinion score (MOS) [12] for all of no-reference (NR), reduced-reference (RR) and full-reference (FR) assessments; the other is the differential mean opinion score (DMOS) [10, 13] for only FR assessment. Recently, there have emerged several subjective VQA methods for other types of videos. For example, Pourashraf *et al.* [14] adopted DMOS of classical subjective VQA method to evaluate the subjective quality of video conferencing. Besides, ITU also extended their DMOS-based VQA method for stereoscopic videos [15], by taking into consideration the characteristics of stereoscopic videos. Nevertheless, there exists few work on subjective VQA of panoramic videos, despite panoramic videos being flooding into our daily life. To our best knowledge, the only work on subjective VQA was presented in [1], in which subjects are enforced to view one region of panoramic videos and then the conventional subjective VQA method of 2D videos is simply applied. However, this is not in accordance with free-view visual experience on panoramic videos.

In this paper, we propose a subjective VQA method to assess quality loss of impaired panoramic videos, in for-

*Corresponding author: Mai Xu (maixu@buaa.edu.cn). This work was supported by the NSFC projects under Grants 61573037 and 61202139, and Fok Ying-Tong education foundation under grant 151061.

978-1-5090-6067-2/17/$31.00 ©2017 IEEE

Table 1: Panoramic video categories of test sequences.

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

(a) HondaF1 (b) RingMan (c) RollerCoaster (d) StarWars (e) Symphony

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

m of DMOS metric[1], complying with human experience on panoramic videos. First, we establish a database containing viewing directions of 40 subjects on watching panoramic videos. Then, by mining our database, we find out that the viewing directions among different subjects are highly consistent with regard to video content. Upon this finding, we develop two subjective VQA metrics, namely Overall-DMOS (O-DMOS) and Vectorized-DMOS (V-DMOS), for rating the overall and regional visual quality reduction of impaired panoramic videos, respectively. Finally, we validate the effectiveness of our VQA method by measuring the Spearman rank correlation coefficient (SRCC) of O-DMOS/V-DMOS values between different groups of subjects. Our contributions in this paper are two-fold:

- We establish a viewing direction database for panoramic videos, with consistency analysis on viewing directions of different subjects.
- We propose a new method for subjective VQA on panoramic videos, taking advantage of our consistency analysis of viewing directions.

2. CONSISTENCY ANALYSIS ON VIEWING PANORAMIC VIDEOS

On account of the omnidirectionality of panoramic videos, 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 videos. Thus, this section mainly discusses the consistency analysis on viewing panoramic videos.

2.1. Database

We establish a new database, which contains viewing direction data of 40 subjects on viewing panoramic videos. In all, there are 48 sequences of panoramic videos in our database. These sequences are diverse in term of their content, and they can be categorized according to video content, as shown in Table 1. All of those 48 sequences were downloaded from YouTube and VRCun. Then, they were cut into short clips, and

[1]Note that method in this paper mainly focuses on FR subjective VQA.

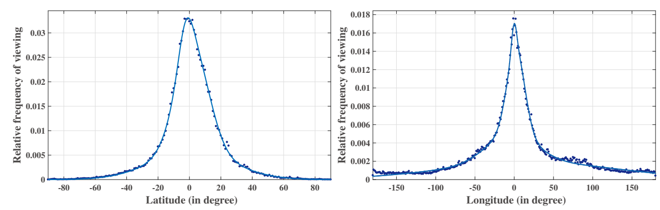


Fig. 2: Viewing direction frequency along with longitude and latitude.

their durations vary from 20 to 60 seconds. The audio tracks were discarded to avoid the impact of acoustic information. To ensure QoE, the resolution of the sequences is beyond 3K (2880×1440) and up to 8K (7680×3840).

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 48 sequences were played at a random order. With the support of the software development kit (SDK) of Vive, we are able to collect the posture data of subjects when viewing panoramic videos. Then, the data of viewing directions to where subject paid attention were obtained in term of Euler angles, and only the inclination and azimuth angles were recorded in our database.

During the experiment, the subjects were seated on a swivel chair, being allowed to turn around freely, such that all regions of panoramic videos are accessible. Besides, to avoid eye fatigue and motion sickness, there was a 5-minute interval after viewing each session of 16 sequences. Our database is available at `https://github.com/Archer-Tatsu/head-tracking`.

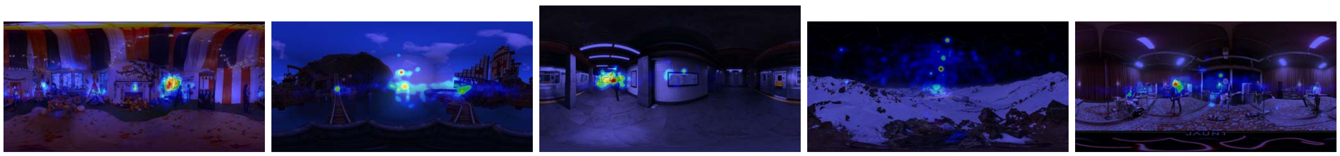
2.2. Data analysis

Now, we analyze the viewing direction data in our database. First, we discard 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 front region. The remaining data are then used for our analysis. Our findings with corresponding analysis are investigated as follows.

Finding 1: When watching panoramic videos, subjects view front region near the equator much more frequent than other regions.

Table 2: CC of each video between the two groups

Category	Name	CC	Category	Name	CC	Category	Name	CC	Category	Name	CC
CA	AcerPredator	0.966	Driving	AirShow	0.971	Others	A380	0.947	Video Game	CS	0.975
	BFG	0.956		DrivingInAlps	0.984		CandyCarnival	0.948		Dota2	0.948
	CMLauncher	0.985		F5Fighter	0.876		MercedesBenz	0.901		GalaxyOnFire	0.970
	Cryogenian	0.932		HondaF1	0.982		RingMan	0.991		LOL	0.949
	LoopUniverse	0.944		Rally	0.965		RioOlympics	0.935		MC	0.947
	Pokemon	0.912		Supercar	0.981		VRBasketball	0.974		SuperMario64	0.969
Movie	Help	0.980	Scenery	Antarctic	0.968	Show	BTSRun	0.979	Action Sports	Gliding	0.880
	IRobot	0.979		BlueWorld	0.949		Graffiti	0.989		Parachuting	0.930
	Predator	0.954		Dubai	0.940		KasabianLive	0.957		RollerCoaster	0.969
	ProjectSoul	0.980		Egypt	0.946		NotBeAloneTonight	0.907		Skiing	0.960
	StarWars	0.995		StarryPolar	0.913		Symphony	0.966		Surfing	0.981
	Terminator	0.971		WesternSichuan	0.958		VRBasketball	0.974		Waterskiing	0.947
In Average		0.956									



(a) CandyCarnival (b) MC (c) Help (d) StarryPolar (e) NotBeAloneTonight

Fig. 3: Examples for a few sequences in which subjects are attracted by other regions.

Figure 1 shows heat maps of viewing directions of some panoramic videos, obtained from all 40 subjects. We can see from this figure that most of viewing directions fall into small regions, located at the front region near the equator. We further statistically calculate viewing directions belonging to different regions of panoramic videos. 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 videos. In this figure, the Gaussian 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 1, which is similar to the conclusion of [5].

Finding 2: In general, there exists high consistency on the viewed regions across different subjects for panoramic videos.

We randomly and equally divide all 40 subjects into two non-overlapping groups, A and B. Then, we generate heat maps of viewing directions of 48 panoramic videos for Groups A and B, which are denoted as $\mathbf{H}_A$ and $\mathbf{H}_B$, respectively. Here, we quantify the correlations of heat maps between $\mathbf{H}_A$ and $\mathbf{H}_B$ using linear correlation coefficient (CC) [16]. Specifically, CC is calculated by

$$\mathrm{CC}(\mathbf{H}_A, \mathbf{H}_B) = \frac{\sum_{x,y}(\mathbf{H}_A(x,y)-\mu(\mathbf{H}_A))\cdot(\mathbf{H}_B(x,y)-\mu(\mathbf{H}_B))}{\sqrt{\sigma(\mathbf{H}_A)^2\cdot\sigma(\mathbf{H}_B)^2}}, \quad (1)$$

where (x, y) is the location of pixel coordinates; $\mu(\cdot)$ and $\sigma(\cdot)$ are mean and standard deviation of the corresponding heat maps. The CC (ranging in $[-1, 1]$) close to $+1$ indicates the high consistency between heat maps $\mathbf{H}_A$ and $\mathbf{H}_B$. Table 2 reports the CC value of each panoramic video between the two groups, which is rather high. It also shows that the CC value averaged over all 48 panoramic videos is 0.956. Thus, it is obvious that subjects behaved consistently when watching panoramic videos. This completes the analysis of Finding 2.

Finding 3: The viewing directions of different subjects are consistent in different regions according to content of panoramic videos, despite being more likely to be attracted by equator and front regions.

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

3. SUBJECTIVE VQA METHOD

In this section, we introduce our subjective VQA method for panoramic videos. In Section 3.1, we present the general arrangements of subjective test for our VQA method. In Section 3.2, the procedure of subjective test is discussed for rating raw quality score of each panoramic video sequence. In Section 3.3, O-DMOS and V-DMOS are proposed as the metric to assess subjective quality of panoramic videos, which are based on their raw scores.

3.1. General arrangements

Panoramic videos are different from 2D videos in playing devices, viewing experience of subjects, etc. Thus, we develop the general arrangements for subjective test on assessing panoramic videos, differing from the test on 2D videos. In the following, we present the general arrangements of the subjective test from the aspects of displaying devices and subjects.

Displaying devices. HMD with its corresponding video player is used to display videos, rather than flat screens on displaying 2D videos. It is because most of panoramic videos are viewed by wearing HMD. In this paper, we use the HTC Vive as the displaying device of HMD and the software VD as the panoramic video player. Besides, VD is also used to project

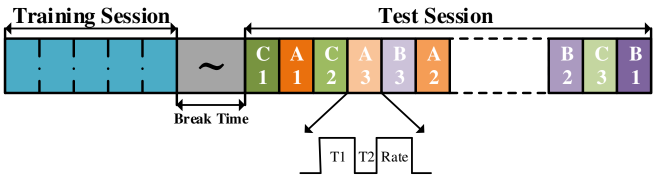


Fig. 4: Structure of the test procedure with two sessions. The SSCQS procedure is illustrated in the part of test session. Ai, Bi and Ci represent various original and impairment sequences from different contents A, B and C, respectively.

our graphic user interface (GUI) for quality rating, allowing the subjects to rate panoramic videos without taking off the HMD. Since panoramic videos can be viewed from different viewing directions, a swivel chair is provided to subjects when viewing panoramic videos.

Subjects. According to Finding 2, the viewing directions of subjects are highly consistent. Therefore, fixing the viewing regions of panoramic videos in [1] is not necessary. Instead, subjects are able to view all content of panoramic videos in our subjective test freely. This way, the rating scores by our method are in accordance with daily visual experience in viewing panoramic videos, in which subjects are free to access all parts of panoramic videos. In addition, the initialization of viewing direction is required when watching panoramic videos, which is different from viewing 2D videos. In our test, the viewing directions of all subjects should be initialized to be the center of front region in panoramic videos, as Finding 1 finds that subjects are more likely to be attracted by this region. However, there still exists slight inconsistency of viewed regions in panoramic videos as analyzed in Finding 2 and 3. Thus, more subjects should be involved in the subjective test for rating quality of panoramic videos, than at least 15 subjects required in [8]. We recommend that at least 20 subjects are required for rating quality scores of panoramic videos, as verified in Section 4.3.

3.2. Test procedure

Training and test. Generally speaking, the test procedure of our subjective VQA method is composed of two sessions: training and test sessions, as shown in Figure 4. The training session is introduced, as some subjects may be unfamiliar with viewing panoramic videos. In the training session, subjects are told about the goal of our test. Then, they need to watch a group of training sequences at different quality, in order to familiarize themselves with panoramic videos and their 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 videos, subjects are more likely to incur eye fatigue and motion sickness. Thus, the maximum duration of test session is 30 minutes [8]. If the test sequences are more than 30 minutes, a short break (at least 3 minutes) with the HMD being taken off needs to be 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 at random order and 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 videos with the same content, which differs from viewing characteristic of 2D videos. After viewing each sequence, subjects are required to rate its quality. The grading scores in the test session are achieved by a continuous-scale slider with a cursor in our quality rating GUI. The score Q has a range of 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 then processed. One is the raw subjective quality scores of the panoramic video sequences as mentioned above. The other is the viewing direction data of subjects during sequence displaying, which make the quality score related to the viewed regions of panoramic videos. This also enables the calculation on V-DMOS, to be discussed next.

3.3. Processing of subjective scores

O-DMOS. Given raw quality scores of each sequence, we follow the DMOS calculation method of 2D videos in [13] to compute O-DMOS, which indicates the overall quality of each panoramic video sequence. Specifically, the difference across quality scores between 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 its corresponding reference sequence. Then, the difference score d_{ij} can be obtained by

$$d_{ij} = S_{ij}^{\text{ref}} - S_{ij}. \quad (2)$$

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

$$\mu_i = \frac{1}{M_i}\sum_{j=1}^{M_i} d_{ij}, \quad \sigma_i = \sqrt{\frac{1}{M_i - 1}\sum_{j=1}^{M_i}(d_{ij} - \mu_i)^2}, \quad (3)$$

$$Z_{ij} = \frac{d_{ij} - \mu_i}{\sigma_i}, \quad (4)$$

where M_i is the number of test sequences viewed by subject i. Then, Z-score Z_{ij} needs to be linearly rescaled to lie in the range of $[0, 100]$:

$$Z_{ij}^{'} = \frac{100(Z_{ij} + 3)}{6}. \quad (5)$$

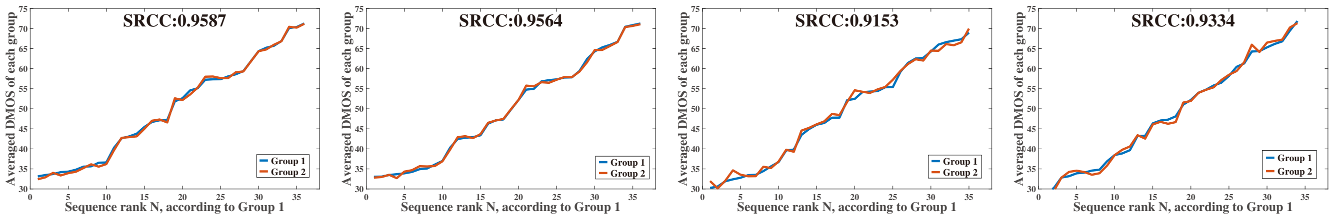
Finally, O-DMOS value of sequence j is computed by averaging $Z_{ij}^{'}$ from N_j valid subjects after subject rejection [8]:

$$\mathrm{O-DMOS}_j = \frac{1}{N_j}\sum_{i=1}^{N_j} Z_{ij}^{'}. \quad (6)$$

V-DMOS. According to Finding 2, there still exists slight inconsistency on viewing directions of panoramic videos. Finding 3 further shows that all regions of panoramic videos are possible to attract human attention. Thus, V-DMOS is used in our subjective VQA method to reflect the quality of

Table 3: The final output in form of V-DMOS of the impaired test sequences. Note that the O-DMOS is included in the V-DMOS as the first element and in bold in the table. The second to the seventh elements in turn represent the DMOS score of the front, left, back, right, top, and bottom region.

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,—]



(a) O-DMOS (b) V-DMOS (Front) (c) V-DMOS (Left) (d) V-DMOS (Right)

Fig. 5: Curves of 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.

different regions of panoramic videos, via making use of the collected raw quality scores and viewing direction data. First, we need to compute the frequency ratio that subject i views region r in sequence j, denoted as f_{ij}^r. Note that f_{ij}^r needs to be normalized for satisfying

$$\sum_r f_{ij}^r = 1. \tag{7}$$

When $f_{ij}^r > f_0$, where f_0 is a threshold, subject i (after subject rejection [8]) 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

$$\text{DMOS}_{jr} = \frac{1}{N_{\mathbf{I}_{jr}}} \sum_{i \in \mathbf{I}_{jr}} Z_{ij}^{'}. \tag{8}$$

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

$$[\text{O}-\text{DMOS}_j \quad \text{DMOS}_{j1} \quad \cdots \quad \text{DMOS}_{jr} \quad \cdots \quad \text{DMOS}_{jR}], \tag{9}$$

where R is total number of regions in panoramic videos. Generally, there exists 6 regions of panoramic videos [17]: front, left, back, right, top, and bottom. As a result, our V-DMOS is able to qualify both overall and regional quality degradation for impaired panoramic videos.

4. EXPERIMENT

4.1. Test benchmark and setting

In this section, we validate the effectiveness of our subjective VQA method by calculating the correlations of O-DMOS and V-DMOS values from two groups of subjects. First, all 12 uncompressed panoramic video sequences of [18] (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 frame rate of 25 fps. Then, H.265 is used to compress these 12 sequences at 3 bit-rates, which are under equirectangular projection. Thus, there are in total 12 reference and 36 impaired sequences.

Then, a total of 48 subjects participated in the subjective test for our VQA method (already presented in Section 3). In the test, subjects were required to view and rate all sequences for raw subjective scores. Next, O-DMOS and V-DMOS are calculated with the rated raw scores. Here, we simply set threshold f_0 to be $1/6$ in V-DMOS calculation, as there are 6 regions in our panoramic videos. Note that no subject is rejected for calculating the O-DMOS and V-DMOS values after using the subject rejection scheme of [8]. Finally, the values of O-DMOS and V-DMOS obtained from raw quality scores of 48 subjects are reported in Table 3.

4.2. Evaluation on effectiveness of our VQA method

Now, the effectiveness of our subjective VQA method is verified by evaluating correlations between 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 correlations of O-DMOS/V-DMOS values between these two groups, averaged over 30 trials, are evaluated. Figure 5 shows the curves of ranked O-DMOS/V-DMOS values[2] for all 36 impaired sequences obtained by group 1, and it also plots the O-DMOS/V-DMOS values by group 2 for the sequences ranked by group 1. We can see from this figure that the correlations of these two groups of O-DMOS/V-DMOS values are extremely high. We quantify such correlations by SRCC, which is also shown in Figure 5. The high SRCC values again indicate the agreement between two groups for O-DMOS and V-DMOS. This implies the effectiveness of our subjective VQA method, as agreement is achieved over dif-

[2]Due to space limitation, we only show the values of front, left and right regions for V-DMOS.

Table 4: SRCC between O-DMOS and V-DMOS of different regions.

V-DMOS region	Front	Left	Right	Back	Top	Bottom
O-DMOS	0.9972	0.9794	0.9750	0.8844	-0.0857	0.8487

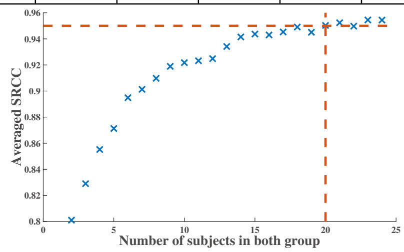


Fig. 6: SRCC of O-DMOS scores between two groups with increasing numbers of subjects in both groups. This is also an averaged result over 30 trials.

ferent tests (two randomly selected groups can be seen as the results from two subjective tests).

4.3. Performance analysis of our VQA method

It is necessary to investigate what is the minimum number of subjects required for our subjective VQA method. To this end, we measure SRCC of O-DMOS values between two groups with different numbers of subjects. Accordingly, Figure 6 shows SRCC along with increased number of subjects in group 1 and 2, which is also the averaged result over 30 trials. We can see that SRCC is converged when subjects are more than 20. Thus, we recommend that the minimum number of subjects for our VQA method should be 20.

It is also interesting to find the relationship between O-DMOS and V-DMOS values of different regions. Table 4 shows SRCCs between O-DMOS and V-DMOS values of different regions, which are calculated from all 48 subjects. It is obvious that the V-DMOS of the front, left and right regions are of strong correlations with the O-DMOS. By contrast, the V-DMOS of the back and bottom regions are generally correlated with the O-DMOS. However, SRCC for the V-DMOS of the top region is rather small. It is because V-DMOS values of the top region are determined by only few subjects, since most of 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 V-DMOS metric.

5. CONCLUSION

In this paper, we have proposed a subjective VQA method for evaluating quality degradation of impaired panoramic videos. Different from conventional subjective VQA methods, our method considers the $360 \times 180°$ FoV of panoramic videos. For such consideration, we conducted the experiment to obtain a new database, which contains viewing directions of 40 subjects on viewing 48 panoramic video sequences. Then, we found from our database that subjects consistently prefer to look at the center of front region of panoramic videos, but there still exists little inconsistency on viewing different regions. In light of our findings, the general arrangements and procedures of the subjective test were developed, such that the quality scores of each panoramic video sequence can be rated by several subjects. Based on the rated quality scores, two metrics, O-DMOS and V-DMOS, were proposed in our subjective VQA method, measuring the overall and regional quality reduction of impaired panoramic videos. Finally, experimental results validate the effectiveness of our subjective VQA method.

6. REFERENCES

[1] Vladyslav Zakharchenko, Kwang Pyo Choi, and Jeong Hoon Park, "Quality metric for spherical panoramic video," in *SPIE Optical Engineering+ Applications*, 2016, pp. 99700C–99700C.

[2] WJ Sarmiento and C Quintero, "Panoramic immersive videos-3d production and visualization framework," in *SIGMAP*, 2009, pp. 173–177.

[3] Robert Konrad, Emily A Cooper, and Gordon Wetzstein, "Novel optical configurations for virtual reality: evaluating user preference and performance with focus-tunable and monovision near-eye displays," in *ACM CHI*, 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] Matt Yu, Haricharan Lakshman, and Bernd Girod, "A framework to evaluate omnidirectional video coding schemes," in *IEEE ISMAR*. IEEE, 2015, pp. 31–36.

[6] Recommendation BT ITU-R, "710-4: Subjective assessment methods for image quality in high-definition television," Tech. Rep., Technical Report, ITU-R, 1998.

[7] ITUT Rec, "P. 910: Subjective video quality assessment methods for multimedia applications," *Int. Telecomm. Union, Geneva*, 2008.

[8] ITUR Rec, "Bt. 500-13,," *Methodology for the subjective assessment of the quality of television pictures, Jan*, 2012.

[9] Margaret H Pinson and Stephen Wolf, "Comparing subjective video quality testing methodologies," in *IEEE VCIP*, 2003, pp. 573–582.

[10] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan C Bovik, and Lawrence K Cormack, "A subjective study to evaluate video quality assessment algorithms," in *IS&T/SPIE Electronic Imaging*, 2010, pp. 75270H–75270H.

[11] 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. 605906–605906.

[12] Thiow Keng Tan, Rajitha Weerakkody, Marta Mrak, Naeem Ramzan, Vittorio Baroncini, Jens-Rainer Ohm, and Gary J Sullivan, "Video quality evaluation methodology and verification testing of hevc compression performance," *IEEE TCSVT*, vol. 26, no. 1, pp. 76–90, 2016.

[13] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K Cormack, "Study of subjective and objective quality assessment of video," *IEEE TIP*, vol. 19, no. 6, pp. 1427–1441, 2010.

[14] Pedram Pourashraf, Farzad Safaei, and Daniel R Franklin, "Minimisation of video downstream bit rate for large scale immersive video conferencing by utilising the perceptual variations of quality," in *IEEE ICME*. IEEE, 2014, pp. 1–6.

[15] IT Union, "Subjective methods for the assessment of stereoscopic 3dtv systems," *Recommendation ITU-R BT*, vol. 2021, 2015.

[16] Jia Li, Changqun Xia, Yafei Song, Shu Fang, and Xiaowu Chen, "A data-driven metric for comprehensive evaluation of saliency models," in *IEEE ICCV*, 2015, pp. 190–198.

[17] Emil Praun and Hugues Hoppe, "Spherical parametrization and remeshing," in *ACM TOG*. ACM, 2003, vol. 22, pp. 340–349.

[18] IEEE1857.9 1st Beijing, "1857.9-01-n0001 output document," 2016.