IVQAD 2017: An Immersive Video Quality Assessment Database

Huiyu Duan, Guangtao Zhai, Xiaokang Yang, Duo Li, Wenhan Zhu
Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai, China
{*hyuduan, liduoee*}*@gmail.com*, {*zhaiguangtao, xkyang, zhuwenhan823*}*@sjtu.edu.cn*

***Abstract*—This paper presents a new database, Immersive Video Quality Assessment Database 2017 (IVQAD 2017), intended for immersive video quality assessment in virtual reality environment. Video quality assessment (VQA) plays an important role in video research fields. Nowadays virtual reality technology have been widely used and playing videos in virtual reality visual system is becoming more and more popular. However, existing research in VQA fields mainly focus on traditional videos. In this paper, we build the IVQAD which contains 10 raw videos and 150 distorted videos. Bit rate, frame rate and resolution were considered as quality degradation factors. All the videos were encoded with MPEG-4. Subjects were asked to assess the video under virtual reality environment and mean opinion score (MOS) was derived by computing. Using IVQAD 2017, researchers can explore the influence of resolution, video compression and video packet loss on immersive videos' quality.**

***Index Terms*—Immersive video, quality assessment, database, virtual reality**

I. Introduction

Quality assessment has long been an important topic in the field of image and video processing. After many years efforts, image quality assessment (IQA) has developed gradually mature. There were many papers discussing IQA methods, such as [1][2][3][4][5][6]. VQA research also developed fast, many researchers have came up with a lot of approaches, such as [7][8][9][10][11]. And research related to IQA or VQA will continue developing in the future.

Video quality assessment (VQA) has been a hot topic for many years and plays an important role in video compression, video processing and video communication fields. With the quick developing of virtual reality technology, the playing platform of videos has gradually transfered from traditional displays to head-mounted displays. And VQA will also step into the virtual reality stage, i.e., Immersive Video Quality Assessment (IVQA). So we established an immersive video quality assessment database (IVQAD) for future quality assessment research. The database consists of three different resolutions with various bit rate and frame rate which has evaluated 10 different scenes. Thus researchers can utilize this database to study a series of quality assessment algorithms of immersive videos.

The main difference between immersive videos and traditional videos in two-dimensional screens is that there is serious distortion at the edge of immersive videos for playing in virtual reality visual system. And the scores of IVQA can only be taken down from dictation of subjects. The main difference between VQA and IVQA is that IVQA experiments conduct with virtual reality head-mounted displays. Therefore, there are many items should be noticed when conducting experiments. Detailed items will be discussed later.

There are already some widely used image quality databases such as LIVE Image Quality Assessment Database [12] and TID 2013 [13]. There are also many video quality assessment databases such as LIVE Video Quality Database [14]. However, these databases evaluated the quality of images or videos only on two-dimensional screens.

With the development of virtual reality technology, research on virtual reality visual system has becoming more and more important. And video quality assessment will be a necessary part of it. There have been much research about traditional video quality assessment so far, nevertheless we still research little about IVQA in virtual reality visual system. Therefore we provided this database for future IVQA, which consists of 160 videos. In our database, there are three different resolutions and the highest resolution is 4096×2048. Immersive videos are taken from widely range of scenes, which makes this database can be widely used in IVQA.

We conducted an experiment to obtain subjective data. Subjective scores collected by plenty of experiments of different subjects are an essential part of video quality assessment. Comparative studies of video quality metrics have examples carried out by VQEG (Video Quality Expert Group). But in our experiment, comparative experiment is hard to conducted since it is proceed under head-mounted diaplays. There are 13 people in our experiment assessing the immersive video. Our goal in this paper is to introduce a new database IVQAD 2017 for the new field IVQA. Using this database, researchers can study VQA of virtual reality visual system for future studies such as video compression, video processing, video communication and so on in virtual reality environment.

The rest of this paper is organised as follows. Section 2 introduces video sources and described detailed experimental procedures. Section 3 analysed subjective scores in the database and evaluated video quality in virtual reality visual system. Conclusion of this paper was in Section 4.

II. Experimental Procedures

A. Shooting and processing

In the database, there are 10 raw videos taken by Insta360 4K Spherical VR Video Camera with resolution of 4096×2048 and file format of MPEG-4, which are fit for playing in virtual reality visual system. Fig. 1 shows the camera we used. And the scenes we saw in virtual reality visual system can cover the whole room. A variety range of scenes of videos are shot in one university campus ranging from lawn to buildings with some people.

Fig. 1. Insta360 4K

The raw data of the videos were exported from the camera at resolution of 4096×2048 and frame rate of 30 fps with audio. Fig. 2 shows sample frames of raw videos shot by Insta360 4K. The typical scenes include lake, bridge, tennis court and so on. In the figure, the brightness varies from one to another. Although they look similar in Fig. 2, huge differences will be seen when watching videos in virtual reality environment. And this will influence the result. So it illustrates the brightness have influence on experience of immersive video. Detailed analysis will be given in Section 3.

To avoid the influence of voice, we eliminate the audio track of all videos using FFMPEG software. Then 10 raw videos were cut into equal length of 15 seconds for fair video quality assessment. To simulate quality degradations, three resolutions were set as 4096×2048, 2048×1024 and 1024×512. And under every resolution, different bit rate and frame rate were set to simulate different bandwidth requirement. The detailed settings are shown in Table I.

TABLE I
FFMPEG Processing Settings Table

type number	Bitrate [kbps]	fps	Resolution
1(raw video)	70000	30fps	4096×2048
2	70000	29.92fps	4096×2048
3	70000	29.92fps	2048×1024
4	70000	29.92fps	1024×512
5	10000	29.92fps	4096×2048
6	10000	29.92fps	2048×1024
7	10000	29.92fps	1024×512
8	2000	29.92fps	4096×2048
9	2000	29.92fps	2048×1024
10	2000	29.92fps	1024×512
11	70000	15fps	4096×2048
12	70000	15fps	2048×1024
13	70000	15fps	1024×512
14	70000	5fps	4096×2048
15	70000	5fps	2048×1024
16	70000	5fps	1024×512

In our database, we only included videos captured by stationary camera from variety of scenes. The situation in which the camera is moving is too complicated. We have shot a few videos with moving camera for preliminary experiment. The result is too bad. Many subjects said the dizziness is unbearable. The experience of watching videos shot by moving camera is bad. And motion sickness in this kind of situation is serious. Although hardware stabilization methods and software stabilization methods were used, experience is still very poor when immersing in virtual reality environment. And subjects are not satisfy with this kind of video.

B. Scoring

To better understand the influence of video resolution, video compression and packet loss on immersive video quality, we have conducted a series of experiments with a group of subjects. And all of the experiments are conducted with the same environment.

The device we used as the virtual reality visual system is HTC Vive which is a kind of virtual reality head-mounted displays. Fig. 3 shows the experiment environment. Since the Virtual reality head-mounted displays is a closed system, we can not print questionnaire for viewer to mark. The scores we got can only coming from testers' oral saying, so it is very time consuming. Nevertheless, to avoid other interferences, we must conducted experiments one by one.

There are many items have been noticed during the experiment. First of all, the sequence of the video streams are randomly displayed in order to ensure that fair scores were obtained. We first conducted an experiment with thirteen subjects using the videos captured by stationary camera of insta360 4K and scored the video with HTC Vive. Each video stream was displayed for 15 seconds. At the end of each test video stream, a 2-4 seconds' voting time were followed. The 5 point ITU continuous scale in the range 0-5 as described in [15] was used when rating the quality of videos in virtual reality environment.

Although the resolution of immersive videos can be 4K, which is 4096×2048 in our experiment. The visual effect is still barely satisfactory in virtual reality environment. However, for better research in the future IVQA, we stretched scores and subject would look through raw immersive videos before starting the experiment. And these videos were told "excellent". Then the disordered video streams would be played.

With a lot of preliminary experiments, we find that 15 seconds is just fit for subjects to look around the entire scenes under virtual reality head-mounted displays. Therefore, the time that subjects used to watch a video must be strictly limited to 15 seconds for fair quality assessment. Additionally, long-time immersion in virtual reality environment will cause dizziness and have influence on immersive video quality assessment. Some researches have raised virtual reality dizziness, such as [16][17][18]. The dizziness existing in virtual reality environment comes mainly from two aspects. For videos captured by stationary camera, it mainly comes from

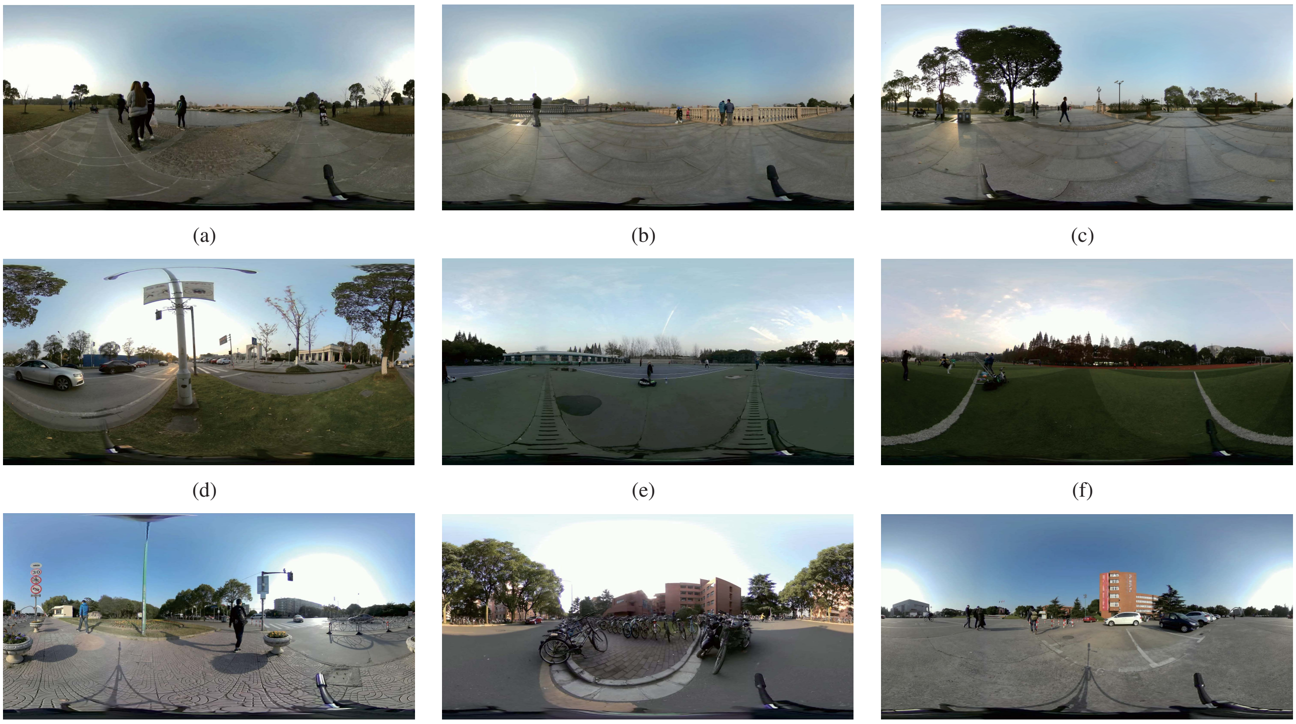

Fig. 2. Sample scenes of the video (a)scene 1 lawn and lake; (b)scene 2 bridge; (c)scene 3 park; (d)scene 4 road; (e)scene 5 tennis court; (f)scene 6 playground; (g)scene 8 sidewalk; (h)scene 9 bicycle; (i)scene 10 square.

tiredness, since the virtual reality head-mounted display is too heavy and field of vision is too narrow. For videos captured by moving camera, it mainly comes from virtual reality motion sickness which sometimes will cause serious dizziness for a long-time. Nevertheless the dizziness in our experiment is slight, we still need to take it into consideration. Since this kind of condition, subjects would take a rest after 8 minutes immersing which is provided by the average discomfort time of subjects.

Fig. 3. Experiment using HTC Vive.

III. Analysis and discussion

A. Analysing data and evaluating immersive video quality

Since the experiment was carried out under virtual reality head-mounted displays, contrast test is difficult and very time-consuming. So the model we used to analyse data is MOS approach.

The Mean opinion Score(MOS) of each videos was computed as:

$$MOS_j = \frac{\sum_{i=1}^{N} m_{ij}^{'}}{N}, \tag{1}$$

where N is the number of subjects and $m_{ij}^{'}$ is the score assigned by subject i to video j with various condition.

The scores assessed by subjects with a grading scale can be of a different type. The grading scale used by subjects can be different types. Five gradations have been used in [19], i.e., "Bad", "Poor", "Fair", "Good", and "Excellent" which correspond to grade "1", "2", "3", "4" and "5". In this paper, we used this kind of grading scale. And MOS was computed under this scale.

Before computing the MOS in IVQAD 2017, there is an important step : Removing outliers. Sometimes, a few subjects will give a score which is far away the mean value. And these outliers should be removed. We used 3σ principle to remove outliers. And σ was computed as follows:

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_{ij}^{\prime} - MOS_j)^2} \tag{2}$$

If a score is beyond 3σ region of MOS, the score will be removed, and MOS will be computed again using the new group of data.

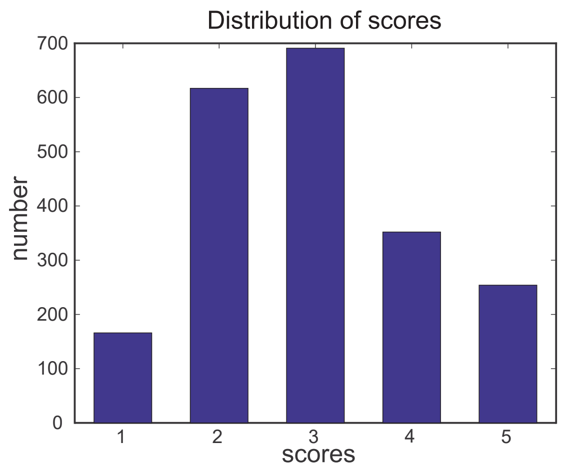

Fig. 4. MOS histogram for IVQA Database.

Fig. 4 shows the MOS histogram for IVQAD 2017. In the result of MOS, the scores mainly centralized among the score “2”, “3” and “4”, and more on “2” and “3”. The numbers of videos from score “1” to “5” are 166, 617, 691, 352 and 254.

All of the MOS were plotted in Fig. 5 with 10 scenes. And MOS of different scenes were expressed with different color. X-axis represents the dispose type of immersive video streams. And “type number” in Table I from top to bottom corresponding to the value 1-16 of x-axis. Y-axis represents MOS of different processing type of 10 scenes. From the figure, we can see that most of the points of one situation gathered. And the MOS of different scenes varied within a limited range. It indicates that under the general trend, immersive video’s quality plays the leading role in IVQA, and in partial analysis, scenes will have a little influence on it.

To analyse IVQA, we choose a sample of MOS of scenes. Fig. 6 shows the MOS of all videos in scene 2 and scene 6 which are shot on a bridge and a playground. The significance of x-axis and y-axis is the same with Fig. 5. It is obvious that there are five group decrease columns in the figure. Five group decreases represent that with the decrease of resolution, the MOS of IVQA gradually reduce. This means the quality of immersive videos is worse and worse as the decrease of resolution. Comparing the scene 2 and scene 6, which were colored by green and yellow separately, the MOS of scene 2 is larger than the MOS of scene 6 in every situation. Scene 2 was shot in the afternoon and scene 6 was shot at dusk. The brightness in scene 2 is larger than it in scene 6. So it illustrates the brightness have influence on the experience of immersive videos.

Fig. 7 comes from scene 2 of Fig. 6, and is listed for better compare. In the Fig. 7, it is obvious that immersive video quality is decrease with the reduce of bit rate and frame rate. Additionally, we can see that the decrease degree of (a) (c) and (e) is slight than that of (b) (d) and (f) since x-axis decrease degree the former is larger than the latter. That is to say, the frame rate has more influence on IVQA than bit rate.

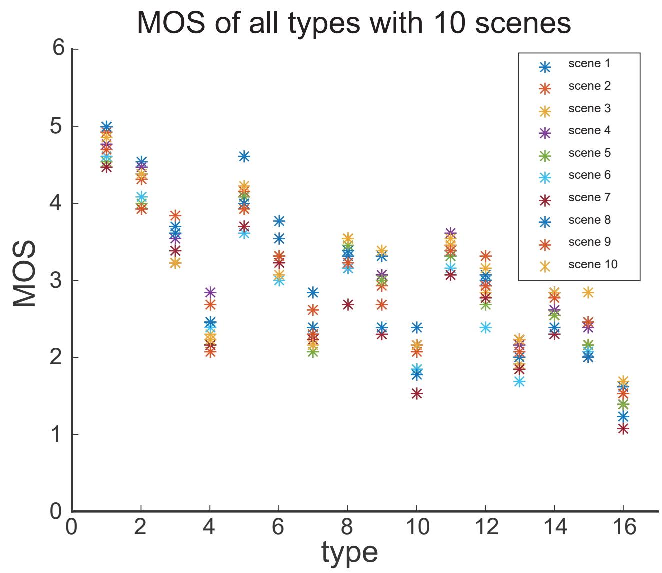

Fig. 5. All MOS in IVQA Database.

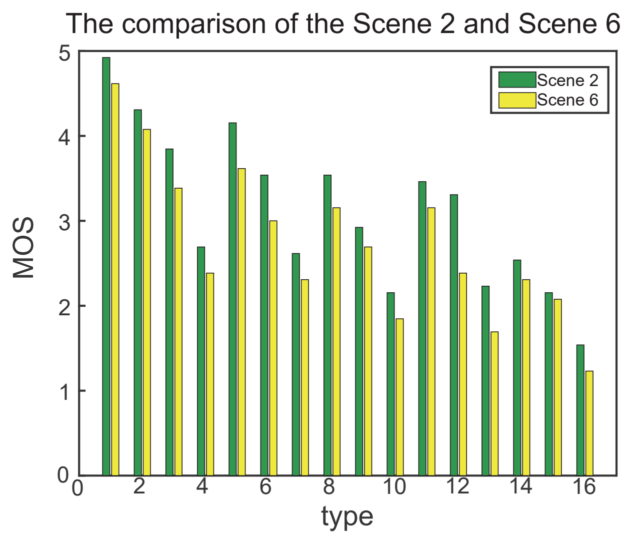

Fig. 6. MOS histogram of scene 2 : bridge.

Combining the Fig. 6, we can get the conclusion that resolution and frame rate have the almost same influence on IVQA more than bit rate. As for the force of influence compared between resolution and frame rate, It varies from person to person in our experiment which comes from asking the question to subjects after experiment. And future research is expected.

There are many factors that will influence IVQA, such as resolution, frame rate, bit rate, even scenes’ bright and dark, background and so on. The strength of influence is varied and resolution and frame rate have more influence on IVQA than others. So promoting camera’s resolution and reducing packet loss will be important aspects in future research. Although

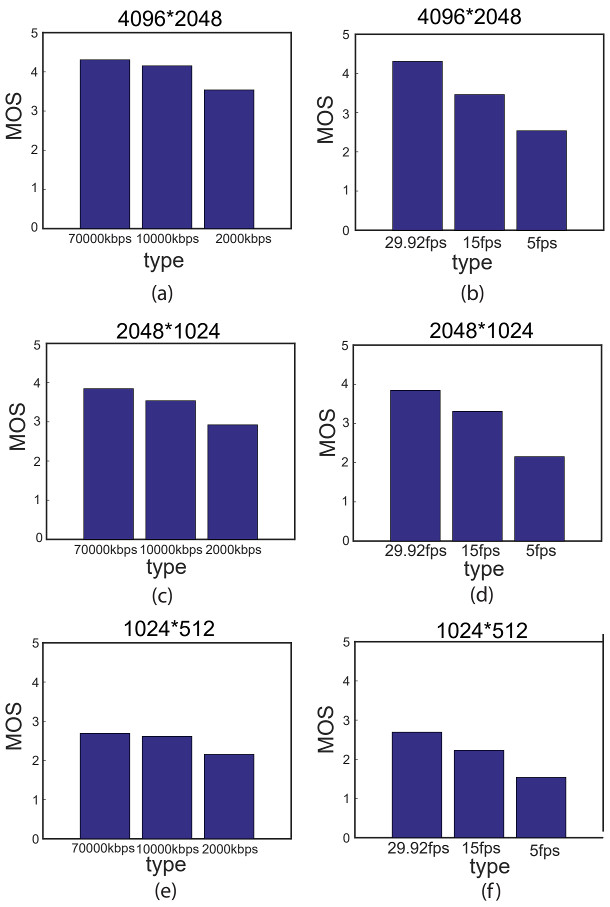

Fig. 7. Comparison of different situation.

videos captured by the moving camera has not been involved, motion sickness is still an important factor of influence on IVQA.

IV. Conclusion

In this paper, we construct a database of Immersive video quality assessment IVQAD 2017. The database contains 160 video streams captured by stationary camera from 10 different scenes and each scene includes a raw video and 15 distorted video. The subjective assessment is conducted with consideration of the watching time, rest time, scoring method and so on. The MOS is derived from thirteen people and then some observation and analysis on the basis of MOS are given. Our goal in this paper is to establish a new database IVQAD 2017 for a new field IVQA. More and more related VQA methods can validate their effectiveness and some new algorithms related to IVQA will be present with the proposal of our database.

Acknowledgment

This work was supported by STCSM (15DZ0500200) and NSFC(61422112, 61371146.

References

[1] Xiongkuo Min, Guangtao Zhai, Ke Gu, Yuming Fang, Xiaokang Yang, Xiaolin Wu, Jiantao Zhou, and Xianming Liu, "Blind quality assessment of compressed images via pseudo structural similarity," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2016, pp. 1–6.

[2] Ke Gu, Weisi Lin, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, and Chang Wen Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Transactions on Cybernetics*, 2016.

[3] Guangtao Zhai, Xiaolin Wu, Xiaokang Yang, Weisi Lin, and Wenjun Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 41–52, 2012.

[4] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang, "Using free energy principle for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 50–63, 2015.

[5] Xiongkuo Min, Guangtao Zhai, Zhongpai Gao, and Ke Gu, "Visual attention data for image quality assessment databases," in *Proc. IEEE Int. Symp. Circuits and Syst.*, 2014, pp. 894–897.

[6] Guangtao Zhai, Wenjun Zhang, Xiaokang Yang, and Yi Xu, "Image quality assessment metrics based on multi-scale edge presentation," in *IEEE Workshop on Signal Processing Systems Design and Implementation, 2005.* IEEE, 2005, pp. 331–336.

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

[8] Guangtao Zhai, Jianfei Cai, Weisi Lin, Xiaokang Yang, Wenjun Zhang, and Minoru Etoh, "Cross-dimensional perceptual quality assessment for low bit-rate videos," *IEEE Transactions on Multimedia*, vol. 10, no. 7, pp. 1316–1324, 2008.

[9] Quan Huynh-Thu and Mohammed Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[10] Yanjiao Chen, Kaishun Wu, and Qian Zhang, "From qos to qoe: a tutorial on video quality assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[11] Yutao Liu, Guangtao Zhai, Debin Zhao, and Xianming Liu, *Frame Rate and Perceptual Quality for HD Video*, Springer International Publishing, 2015.

[12] Hamid R Sheikh, Zhou Wang, Lawrence Cormack, and Alan C Bovik, "Live image quality assessment database release 2," 2005.

[13] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., "Image database tid2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

[14] K Seshadrinathan, R Soundararajan, AC Bovik, and LK Cormack, "Live video quality database," *Laboratory for Image and Video Engineering*, 2009.

[15] P ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," 1999.

[16] EC Regan and KR Price, "The frequency of occurrence and severity of side-effects of immersion virtual reality.," *Aviation, Space, and Environmental Medicine*, 1994.

[17] Robert S Kennedy, Norman E Lane, Michael G Lilienthal, Kevin S Berbaum, and Lawrence J Hettinger, "Profile analysis of simulator sickness symptoms: Application to virtual environment systems," *Presence: Teleoperators & Virtual Environments*, vol. 1, no. 3, pp. 295–301, 1992.

[18] Wei Chen, Jian-Gang Chao, Xue-Wen Chen, Jin-Kun Wang, and Cheng Tan, "Quantitative orientation preference and susceptibility to space motion sickness simulated in a virtual reality environment," *Brain research bulletin*, vol. 113, pp. 17–26, 2015.

[19] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on image processing*, vol. 15, no. 11, pp. 3440–3451, 2006.