Subjective Quality Assessment of Stereoscopic Omnidirectional Image ⋆

Jiahua Xu∗, Chaoyi Lin∗, Wei Zhou, and Zhibo Chen(✉)

CAS Key Laboratory of Technology in Geo-spatial Information Processing and
Application System
University of Science and Technology of China, Hefei 230027, China
`chenzhibo@ustc.edu.cn`

Abstract. Stereoscopic omnidirectional images are eye-catching because they can provide realistic and immersive experience. Due to the extra depth perception provided by stereoscopic omnidirectional images, it is desirable and urgent to evaluate the overall quality of experience (QoE) of these images, including image quality, depth perception, and so on. However, most existing studies are based on 2D omnidirectional images and only image quality is taken into account. In this paper, we establish the very first **S**tereoscopic **O**mnidirectiona**L I**mage quality assessment **D**atabase (SOLID). Three subjective evaluating factors are considered in our database, namely image quality, depth perception, and overall QoE. Additionally, the relationship among these three factors is investigated. Finally, several well-known image quality assessment (IQA) metrics are tested on our SOLID database. Experimental results demonstrate that the objective overall QoE assessment is more challenging compared to IQA in terms of stereoscopic omnidirectional images. We believe that our database and findings will provide useful insights in the development of the QoE assessment for stereoscopic omnidirectional images.

Keywords: Stereoscopic omnidirectional images · Quality of experience · Subjective quality evaluation.

1 Introduction

Virtual reality (VR) has become an eye-catching topic in recent years due to the rapid development of industry ecosystem and technology. The omnidirectional images and videos in VR can provide viewers with immersive experience where viewers can explore every viewing direction freely by wearing the head mounted display (HMD). When Facebook Surround 360, a high-quality 3D-360 video capture system, becomes available in 2016, it is possible to capture and render stereoscopic omnidirectional images. However, owing to the limitation of photographic apparatus, transmission bandwidth, display devices, etc., the overall quality of experience (QoE) of stereoscopic omnidirectional contents is

---

⋆ The first two authors made equal contributions to this work.

(a) Exhibition   (b) Road   (c) Museum

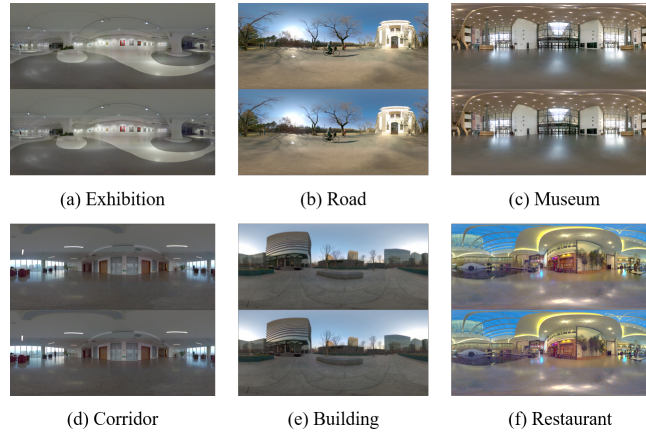(d) Corridor   (e) Building   (f) Restaurant

Fig. 1: Test image contents used in our database.

far from satisfactory. Its worth noting that the overall QoE considers both the factors of image quality and depth perception in this paper. Thus, in order to generate high-quality and realistic stereoscopic omnidirectional images, it is desirable and urgent to evaluate the image quality, depth perception, and overall QoE of these images.

Most existing subjective quality assessment databases are based on 2D omnidirectional contents and only image quality is taken into account. An omnidirectional image quality assessment (OIQA) database [3] has been built for subjective quality evaluation study. Xu et al. [11] propose a subjective visual quality assessment method for omnidirectional videos. Zhou et al. [15] explore the impact of spatial resolution on the perceptual quality of immersive 360-degree images. IVQAD 2017 [4] is a database built for immersive video quality assessment.

Different from 2D omnidirectional images, the overall QoE of stereoscopic omnidirectional images concerns multiple aspects due to the extra dimensionality of immersive contents. Specially, the additional dimension of depth may implicitly affect the experience of viewing. Thus the image quality and depth perception are two important factors which should be taken into consideration when evaluating the subjective perception of QoE. However, to the best of our knowledge, there is no available stereoscopic omnidirectional image quality assessment (IQA) database for investigating the property of these QoE factors.

In this paper, we establish the very first **S**tereoscopic **O**mnidirectiona**L** **I**mage quality assessment **D**atabase (SOLID) by evaluating the image quality, depth perception, and overall QoE of stereoscopic omnidirectional images. The relationships among these three factors are analyzed in our database. We find that the subjective rating scores of overall QoE are highly correlated with image quality and it is moderately correlated with depth perception. Besides, the subjective ratings of image quality are also correlated with depth perception scores.
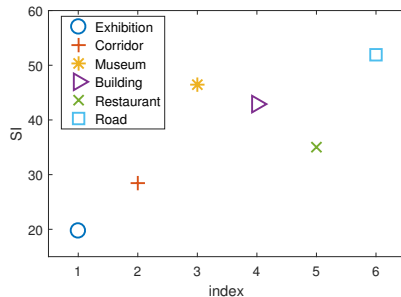
Fig. 2: Spatial information (SI) distrubution for the test images used in our database.

Finally, several well-known objective IQA metrics are tested on our database. Although these classic metrics can achieve a promising performance on predicting image quality, it is still a challenge to predict the overall QoE for stereoscopic omnidirectional images.

The rest of the paper is organized as follows. Section 2 describes the details of our SOLID database. Section 3 analyzes the subjective rating scores to investigate the relationship of multidimensional rating scores and finally several objective IQA metrics are evaluated. Section 4 concludes the paper.

2   Stereoscopic Omnidirectional Image Quality Assessment Database

This section introduces the experiment of subjective quality evaluation for stereoscopic omnidirectional images, including the dataset and subjective test methods.

2.1   Image Database

The test images used in our experiment are captured by Facebook Surround 360 which is an open source hardware and software for generating stereoscopic omnidirectional images and videos. It can render the stereoscopic omnidirectional images (equirectangular projection) with the resolution of 8192 × 8192 and the file format of PNG. Besides, the disparity can be adjusted in the software of Surround 360 before rendering stereoscopic omnidirectional images, which is used to render images with different depth perception levels in our database.

There are 6 high-quality reference stereoscopic omnidirectional images with the resolution of 8192 × 8192 in our database, which are shown in Fig. 1. The stereoscopic omnidirectional images are in the format of top-and-bottom that left and right view images are packed vertically and there exists disparity between left view and right view. It can be seen that there are both indoor scenes and outdoor

Table 1: Distortion level and depth level settings for test images in our database.

HRCID	QP for left view	QP for right view	Depth level
HRC01	-	-	zero disparity
HRC02	25	25	zero disparity
HRC03	40	40	zero disparity
HRC04	48	48	zero disparity
HRC05	25	45	zero disparity
HRC06	25	48	zero disparity
HRC07	45	35	zero disparity
HRC08	51	40	zero disparity
HRC09	-	-	medium disparity
HRC10	25	25	medium disparity
HRC11	40	40	medium disparity
HRC12	48	48	medium disparity
HRC13	25	45	medium disparity
HRC14	51	25	medium disparity
HRC15	45	35	medium disparity
HRC16	35	51	medium disparity
HRC17	45	51	medium disparity
HRC18	-	-	large disparity
HRC19	25	25	large disparity
HRC20	40	40	large disparity
HRC21	48	48	large disparity
HRC22	25	45	large disparity
HRC23	51	25	large disparity
HRC24	45	35	large disparity
HRC25	35	51	large disparity
HRC26	45	51	large disparity

scenes in our database. Besides, the spatial information (SI) which represents the spatial complexity of an image is taken into account when selecting the reference images. Fig. 2 shows SI of the six reference images used in our experiment.

To investigate the relationship between image quality, depth perception, and overall QoE, the stimuli are generated to cover a wide range of image quality and depth perception. The distortion levels for each reference stereoscopic omnidirectional image are presented in Table 1. The hypothetic reference circuits (HRCs) are used to represent the test stereoscopic omnidirectional image with certain distortion and disparity. As shown in Table 1, there are 26 test images for each reference stereoscopic omnidirectional image. Thus, we have 156 test stereoscopic omnidirectional images in our database generated from 6 reference stereoscopic omnidirectional images.

As inspired by [16], there are 3 depth levels for each reference stereoscopic omnidirectional image in our experiment : 1) zero disparity images where there is no disparity between left and right view images, 2) medium disparity images, and 3) large disparity images.

To simulate the quality degradation, each reference image is compressed into Better Portable Graphics (BPG) format [1] with different quantization parameters (QPs). BPG is a new image format which aims to replace the JPEG image format when quality or file size is an issue. Considering the stereoscopic omnidirectional images are usually large in file size, we believe BPG format may be popular in this kind of contents and thus we choose BPG compression distortion in our experiment. The reference stereoscopic omnidirectional images are
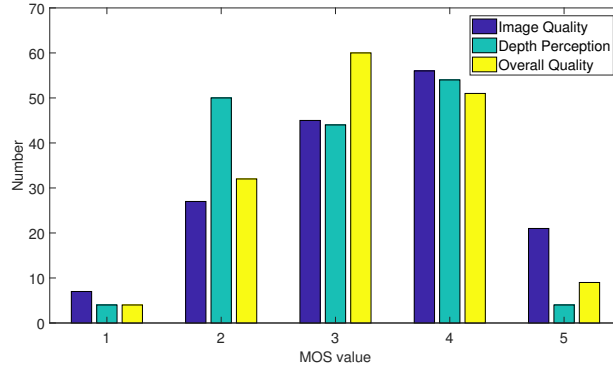
Fig. 3: MOS histogram for image quality, depth perception, and overall QoE in our database.

distorted either symmetrically or asymmetrically in our database. The symmetrical and asymmetrical distortion is determined according to whether the left and right view images are distorted with the same distortion level.

2.2 Subjective Test Methods

In our experiment, image quality, depth perception, and overall QoE are evaluated by the subjects. The experiment is performed according to Absolute Category Rating with Hidden Reference (ACR-HR) which is described in [5]. ACH-HR is a single stimulus evaluation and voting is performed after each viewing. Images are assessed using the five-grade scale with following levels: "5 - Excellent", "4 - Good", "3 - Fair", "2 - Poor", and "1 - Bad".

The following equipment is used in our experiment. The Samsung Gear VR, which is a kind of head-mounted display, is used as the virtual reality display system. The field of view provided by Gear VR is 101 degrees. Samsung Galaxy $S9+$ is used to display the images with a resolution of 2560×1440.

18 non-expert student subjects aged from 21 to 28 years old take part in our subjective test. All participants pass the visual acuity, color vision (Ishihara charts) and stereo acuity tests (RANDOT test). Before the formal test, there is a training session that each subject is explained the purpose of the evaluation. Also, they are shown the examples of different levels of compression artifacts and depth perception. In the formal test, there are 156 images including reference and distorted images to be rated randomly. There is no time limitation for subjects when they are watching stereoscopic omnidirectional images. During the experiment, subjects can take a break as long as they feel tired or uncomfortable to avoid eye fatigue.

Table 2: 95% confidence interval (CI95) of the image quality, depth perception, and overall QoE.

	image quality	depth perception	overall QoE
CI95	0.34	0.45	0.39

Table 3: Correlation between image quality, depth perception, and overall QoE.

	image quality	depth perception
overall QoE	0.90	0.67

3   Analysis of Subjective Database

This section provides a detailed analysis of the subjective evaluating scores. First, the suitability of evaluation methods is analyzed to ensure that data collected from subjects is effective. Second, we explore the relationship between image quality, depth perception, and overall QoE. Finally, performance evaluation of some objective metrics is performed on our database.

3.1   Data Analysis

The subjects whose correlation coefficient with average image quality is lower than 0.8 or with average depth perception is lower than 0.65 are considered as outliers and their subjective evaluating scores are removed from our database. There remain 16 valid subjects (7 males and 9 females) after outlier removal.

Mean Opinion Score (MOS) values are computed for each test image in the database by averaging scores of valid subjects. According to ITU-T P.910 [5], the 95% confidence interval (CI) of the subjective rating scores for image quality, depth perception, and overall QoE are given in Table 2. Results show that all the subjects reach a reasonable agreement on the perceived image quality, depth perception, and overall QoE. Fig. 3 shows the MOS histogram for image quality, depth perception and overall QoE in our database, respectively. In Fig. 3, the MOS values mainly centralize among score 2, 3, and 4.

The MOS values of 6 scenes for 26 HRCs are averaged and shown in Fig. 4. We can draw some conclusions from Fig. 4. First, the BPG compression will affect image quality and overall QoE greatly. Second, depth perception is dominated by the disparity and the BPG compression has a moderate influence on depth perception. Third, overall QoE is affected by image quality and depth perception jointly. The overall QoE scores are consistent with the image quality scores and the strong depth perception tends to enhance the overall QoE while the poor depth perception will reduce the overall QoE. This can be observed from the MOS values of HRCs in Fig. 4, such as HRC 1 and 26. The linear correlation coefficients between image quality, depth perception, and overall QoE are presented in Table 3, which demonstrates that image quality is the dominant factor for overall QoE compared with depth perception.
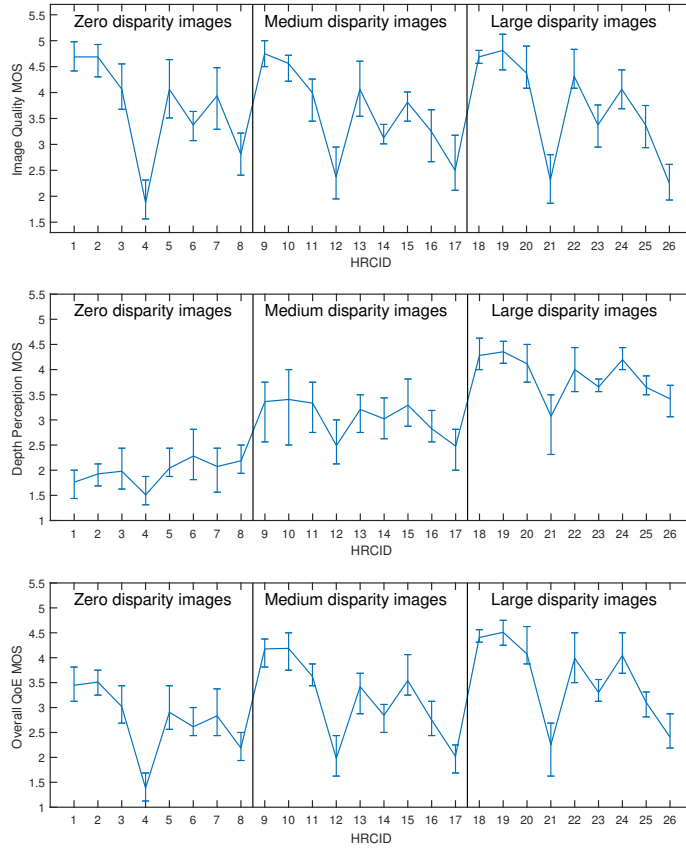
Fig. 4: MOS of image quality, depth perception, and overall QoE for each HRC.

3.2   Key Factors to Image Quality and Depth Perception

The relationship between image quality and distortion levels is shown in Fig. 5. We find that under the same distortion level, images with large disparity tend to have higher image quality scores and this is especially obvious in asymmetrical distortion QP set (Fig. 5 (b)). According to the questionnaire from the subjects, stereoscopic omnidirectional images with strong depth perception can provide a more realistic environment compared to those with weak depth perception. As a result, subjects tend to rate a high score for image quality of these images.

Fig. 6 is the scatter plot of subjective rating scores of image quality vs. that of depth perception. It can be observed that subjective image quality scores are highly correlated with depth perception scores. This is an interesting finding because image quality and depth perception are usually considered as two different perceptual dimensions. Two possible explanations are found from the data analysis and questionnaire. First, image quality degradation will result in blurry
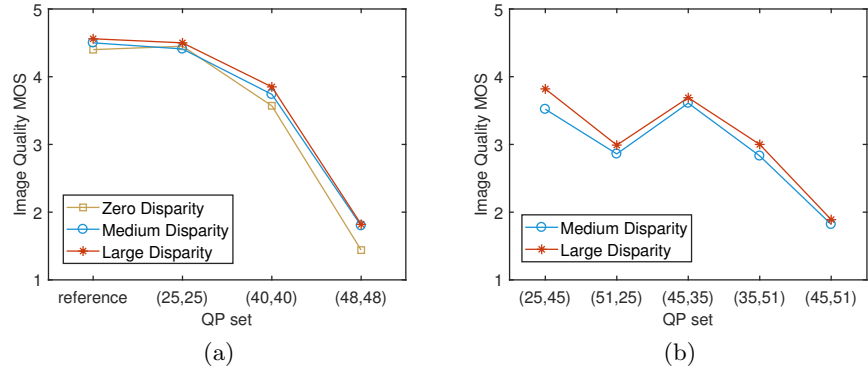
Fig. 5: The influence of distortion levels to image quality. (a) Image quality of symmetrically compressed images at different depth perception levels; (b) Image quality of asymmetrically compressed image at different depth perception levels.
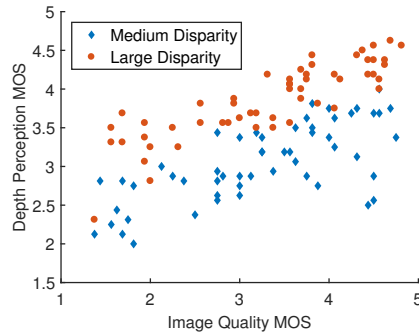

Fig. 6: Relationship between image quality MOS and depth perception MOS.

objects, which will weaken the depth perception. Second, subjects tend to give high images quality scores when depth perception of that image is strong, and vice versa [8]. Thus image quality is highly correlated with depth perception.

3.3   Performance Evaluation of Objective Metrics

We tested nine well-known objective image quality assessment metrics on our database including six 2D IQA metrics PSNR,SSIM [10], MSSSIM [9], FSIM [14], VSI [13] and BRISQUE [6], two 2D omnidirectional IQA metrics S-PSNR [12] and WS-PSNR [7], and a 3D IQA metric [2]. For these 2D metrics, the predicted image quality of left and right view images are averaged as the final quality of stereoscopic omnidirectional images. Linear correlation coefficient (LCC) and Spearmans rank correlation coefficient (SROCC) are used to measure

Table 4: LCC and SROCC performance of objective image quality metrics on our database.

Type	Metrics	Image Quality		Overall QoE	
		LCC	SROCC	LCC	SROCC
2D IQA	PSNR	0.7397	0.6529	0.6295	0.5762
	SSIM [10]	0.8621	0.8742	0.7388	0.7508
	MS-SSIM [9]	0.7839	0.7165	0.6565	0.6147
	FSIM [14]	**0.8922**	**0.8959**	**0.7750**	**0.7742**
	VSI [13]	0.8852	0.8757	0.7623	0.7664
	BRISQUE [6]	0.7496	0.7288	0.6581	0.6550
2D OIQA	S-PSNR [12]	0.7366	0.6423	0.6196	0.5656
	WS-PSNR [7]	0.7238	0.6388	0.6160	0.5639
3D IQA	CHEN [2]	0.7430	0.7037	0.5527	0.5342

the performance of objective metrics. The higher correlation coefficient means better correlation with human subjective quality judgement.

The performance of the above metrics is shown in Table 4 and we can find that FSIM has the best performance while 2D OIQA metrics do not show thier advantages on omnidirectional images compared with traditional 2D IQA metrics. Existing metrics can achieve promising performance for image quality, however the performance drops significantly when it refers to overall QoE. Results in Table 4 demonstrate that predicting overall QoE is still a challenge since it considers not only the image quality but also other factors, such as depth perception, visual comfort, etc.

4   Conclusion

In this paper, we build a **S**tereoscopic **O**mnidirectiona**L** **I**mage quality assessment **D**atabase named as SOLID. There are 156 test images with different levels of depth and BPG compression artifacts in SOLID. To the best of our knowledge, it is the very first work on stereoscopic omnidirectional image quality assessment database, which considers the factors of image quality, depth perception, and overall QoE. By analyzing the subjective rating scores, we find that image quality is the dominant factor for overall QoE compared to depth perception. Besides, image quality is also highly correlated with depth perception that images with large disparity tend to achieve higher image quality scores. Finally, some existing well-known objective IQA metrics are tested on our SOLID database. Experimental results show that overall QoE assessment is more challenging compared to IQA in terms of stereoscopic omnidirectional images.

References

1. Bellard, F.: Bpg image format (http://bellard. org/bpg/) (2017)
2. Chen, M.J., Su, C.C., Kwon, D.K., Cormack, L.K., Bovik, A.C.: Full-reference quality assessment of stereopairs accounting for rivalry. Signal Processing: Image Communication **28**(9), 1143 – 1155 (2013)
3. Duan, H., Zhai, G., Min, X., Zhu, Y., Fang, Y., Yang, X.: Perceptual quality assessment of omnidirectional images. In: 2018 IEEE International Symposium on Circuits and Systems (ISCAS). pp. 1–5 (May 2018)
4. Duan, H., Zhai, G., Yang, X., Li, D., Zhu, W.: Ivqad 2017: An immersive video quality assessment database. In: 2017 International Conference on Systems, Signals and Image Processing (IWSSIP). pp. 1–5 (May 2017)
5. ITU-T RECOMMENDATION, P.: Subjective video quality assessment methods for multimedia applications (1999)
6. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing **21**(12), 4695–4708 (Dec 2012)
7. Sun, Y., Lu, A., Yu, L.: Weighted-to-spherically-uniform quality evaluation for omnidirectional video. IEEE Signal Processing Letters **24**(9), 1408–1412 (Sept 2017)
8. Wang, J., Wang, S., Ma, K., Wang, Z.: Perceptual depth quality in distorted stereoscopic images. IEEE Transactions on Image Processing **26**(3), 1202–1215 (2017)
9. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003. vol. 2, pp. 1398–1402 Vol.2 (Nov 2003)
10. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (April 2004)
11. Xu, M., Li, C., Liu, Y., Deng, X., Lu, J.: A subjective visual quality assessment method of panoramic videos. In: 2017 IEEE International Conference on Multimedia and Expo (ICME). pp. 517–522 (July 2017)
12. Yu, M., Lakshman, H., Girod, B.: A framework to evaluate omnidirectional video coding schemes. In: 2015 IEEE International Symposium on Mixed and Augmented Reality. pp. 31–36 (Sept 2015)
13. Zhang, L., Shen, Y., Li, H.: Vsi: A visual saliency-induced index for perceptual image quality assessment. IEEE Transactions on Image Processing **23**(10), 4270–4281 (Oct 2014)
14. Zhang, L., Zhang, L., Mou, X., Zhang, D.: Fsim: A feature similarity index for image quality assessment. IEEE Transactions on Image Processing **20**(8), 2378–2386 (Aug 2011)
15. Zhou, R., Huang, M., Tan, S., Zhang, L., Chen, D., Wu, J., Yue, T., Cao, X., Ma, Z.: Modeling the impact of spatial resolutions on perceptual quality of immersive image/video. In: 3D Imaging (IC3D), 2016 International Conference on. pp. 1–6. IEEE (2016)
16. Zhou, W., Liao, N., Chen, Z., Li, W.: 3d-hevc visual quality assessment: Database and bitstream model. In: Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on. pp. 1–6. IEEE (2016)