Subjective Video Quality Database for Virtual Reality

Zhenzhong Chen and Yingxue Zhang

Introduction

We establish a subjective video quality database for virtual reality. 48 panoramic video sequences with different levels of compression impairments are viewed and rated through HTC VIVE by 30 non-expert subjects.

With the development of virtual reality (VR) and related technologies, the establishment of immersion calls for higher quality of video contents. However, the processing such as stitching and compressing on the videos greatly influences the quality. Therefore, quality assessment for panoramic videos attaches much importance in specifying and promoting the quality of experience (QoE).

A subjective quality database for panoramic videos is established through a subjective rating test with virtual reality HMD, from which we can 1) figure out the observers' psychophysical response to the VR contents, 2) provide reliable reference for evaluating the performance of the objective assessment methods.

Subjective Quality Assessment Test

As shown in Figure 1 and Table 1, 10 panoramic common test sequences released by MPEG [1] are adopted as reference sequences. All the sequences are in the format of equirectangular (ERP), lasting for 10s each.

The original sequences of different resolutions are sampled to a consistent resolution for presenting on HTC VIVE before coding. Fixed QP values guarantee a consistent quality of different videos on the same compression level.

Coding impairments are introduced to the reference videos to obtain test sequences using HM-16.14 with 360-Lib at 5 QP points, i.e., 22, 27, 32, 37, 42. After processing, a total of 60 sequences on different but relatively stable quality levels are prepared for the experiment, among which reference "AerialCity" and its corresponding impaired sequences are used for training, three sequences from "ChairLift" for stabilizing and the remaining 48 for testing.

Table 1. Information of original test sequences in ERP format [1]

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

The videos are presented one at a time with HTC VIVE and are voted independently. The subjects can view the contents on all directions freely. The reference sequences are also displayed and voted without any special identification, dubbed Hidden Reference [2]. All the test sequences will be presented randomly and only once. Absolute five-grade scale is used to rate the video quality considering the quality range. The final rating scores for the test sequences are defined using Difference Mean Opinion Score (DMOS).

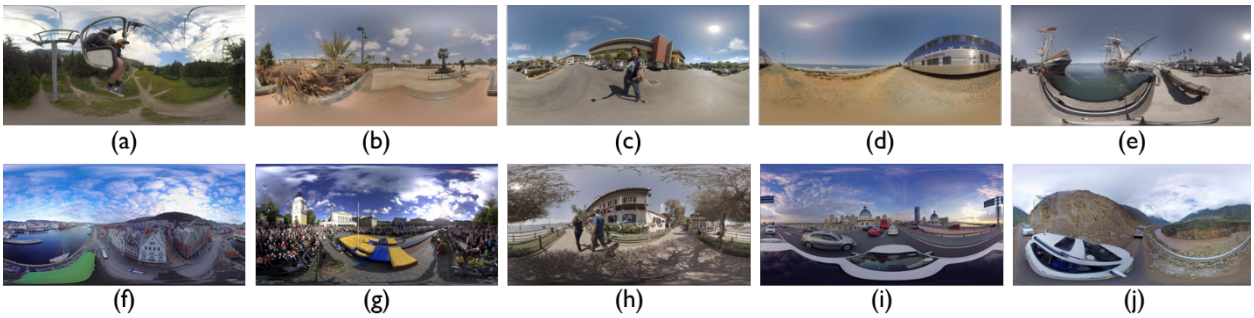

Figure 1. Thumbnails of the ten references used in the test. (a) ChairLift, (b) SkateboardingTrick_le, (c) SkateboardInLot, (d) Train_le, (e) Harbor, (f) AerialCity, (g) PoleVault_le, (h) KiteFlite, (i) DrivingInCity, (j) DrivingInCountry.

As aforementioned, the subjects can view the video freely. Despite of the high consistency on viewing pattern, the free-viewing task will unavoidably lead to some extreme conditions that some subjects may focus on totally different factors from the others. Therefore, the number of subjects for each test is suggested to be more than 15 being recommended for 2D video assessment. A larger number of subjects guarantees the reliability when some extreme data exists.

We recruit 30 subjects to participate in the assessment tests. The subjects are undergraduate and graduate students, including 17 males and 13 females. None of the subjects majors in quality assessment or related areas, nor do they involve in the design or further analysis of the tests. They are asked to evaluate the overall quality of the video.

The Subjective Video Quality Database

After experiment, a set of rating data is obtained for all the sequences. Before calculating DMOS, post-experiment screening is conducted to assess subject reliability and ensure a valid data set. If a subject does not respond according to the instructions, the data has to be discarded. Firstly, a subject's data will be discarded if there is any missed rating [3]. Secondly, the subject with unreliable ratings will also be screened, which is specified in [4].

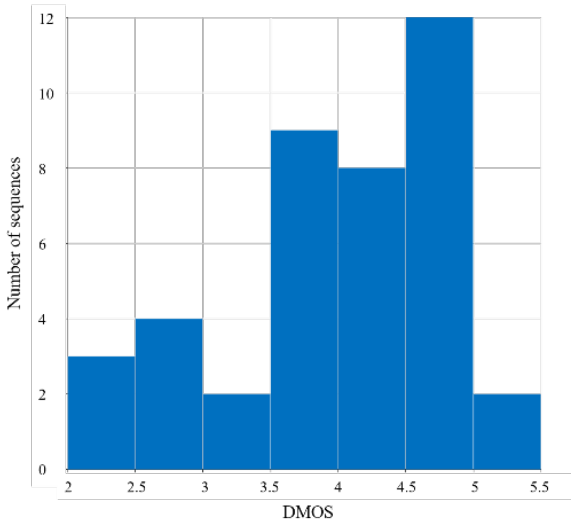

Figure 2. Histogram of the DMOS uniformly spaced between the minimal and maximal values.

In total, the ratings of 3 subjects are discarded by the screening process. DMOS is calculated with the remaining reliable scores on reference and test sequences.

Figure 2 shows the histogram of the DMOS for all the test sequences. The DMOS lies in the range of [2.04, 5.08], corresponding to mean Z-score range of [-0.6, 3], which covers approximately 72% of the area of standard normal distribution.

Conclusion

A subjective quality database for panoramic videos is established through a compact subjective rating test involving 30 subjects and 60 sequences with coding impairments of different levels. The DMOS of the sequences is calculated on the basis of validated subjective ratings and is reasonably distributed on the quality range. Therefore, the database can be promising in further VR applications.

References

[1] J. Boyce, E. Alshina, A. Abbas, Y. Ye, "JVET common test conditions and evaluation procedures for 360° video", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D1030, 4th Meeting, Oct. 2016.

[2] ITU-T, "Subjective video quality assessment methods for multimedia applications", Rec. ITU-T P. 910, Apr. 2008.

[3] VQEG, "Final Report from the video quality experts group on the validation of objective models of video quality assessment", Mar. 2000.

[4] ITU-R, "Methodology for the subjective assessment of the quality of television pictures", Rec. ITU-R BT. 500, Jan. 2012.

Zhenzhong Chen *is a professor at Wuhan University. He leads the Institute of Intelligent Sensing and Computing conducting researches on multimedia technology, computer vision, image processing and understanding, data mining for geoinformatics, etc.*

Yingxue Zhang *is a Ph. D. student at Wuhan University. She is a member of the Institute of Intelligent Sensing and Computing. Her research interests include quality assessment and computational vision.*