國立清華大學電機資訊學院資訊系統與應用研究所

碩士論文

Institute of Information Systems and Applications
College of Electrical Engineering and Computer Science
National Tsing Hua University

Master Thesis

# 360度全景影片於頭戴虛擬實境之使用者體驗品質建模

# Modeling Quality-of-Experience of 360° Videos in Head-Mounted Virtual Reality

姚舜懷
Shun-Huai Yao

學號：105065507
Student ID:105065507

指導教授：徐正炘 博士
Advisor: Cheng-Hsin Hsu, Ph.D.

中華民國 107 年 12 月
December, 2018

國立清華大學
資訊系統與應用研究所

碩士論文

# 360度全景影片於頭戴虛擬實境之使用者體驗品質建模

姚舜懷 撰

107
12

# 致謝

首先，我要特別感謝我的指導教授：徐正炘教授，不管是研究或職涯課題，他總是給予我很多的指導以及建議，在過去的學習過程中，我從我的指導教授身上學到了很多。我也要感謝網路與多媒體系統實驗室的同學們，特別是樊慶玲，在研究的過程中，她幫助了我很多，如果沒有她的幫助，我一定沒辦法順利畢業以及完成如此多的事情。在此我也要感謝在過去碩士生涯中所有幫助過我的人以及曾經願意前來做實驗的受測者，如果沒有你們，我一定沒有辦法順利完成我的論文。最後，我想感謝洪華駿，雖然我們的研究領域不同，但他依舊幫助了我很多，包含電腦網路的知識以及溝通技巧。

# Acknowledgments

First, I would like to thank my advisor: Dr. Cheng-Hsin Hsu. He consistently gave me a lot of guidance and suggestions in many aspects, including research and career. Through him, I learned a lot. Secondly, I would like to thank my colleagues in Networking and Multimedia Systems Lab, especially Ching-Ling Fan. Without Ching-Ling's careful tutoring, I would not be able to graduate and finish many tasks. I would like to express my gratitude toward all the people who helped me and all the subjects whom I recruited for my experiments. I would not be able to complete my thesis without your help. Lastly, I would like to thank Hua-Jun Hong. Although we had different research topics, he helped me a lot in many aspects, such as computer networking knowledge and communication skills.

# 中文摘要

配戴頭戴顯示器觀看360度全景影片愈趨流行，然而在這領域理解使用者體驗品質(QoE)是一項很大的挑戰，因為使用者體驗被諸多因素所影響。本論文致力於提出使用者體驗品質模型以預測配戴頭戴顯示器觀看360度全景影片的使用者體驗。為了達到此目標，我們實作了一360度全景影片播放器，其支援多種不同的投影方式，此影片播放器可支援4K解析度和每秒30幀數的360度全景影片。為了探索會影響使用者體驗的顯著因素，我們設計一使用者實驗，並考慮以下因素：（一）投影方式、（二）壓縮量化參數及（三）影片特質，我們分析其因素與使用者體驗之間的相關程度。我們更根據以上實驗結果提出了多個使用者體驗品質模型，以預測使用者的360度全景影片觀看體驗，我們也針對不同使用情境推薦了其適合的模型。

# Abstract

Watching 360° videos with a Head-Mounted Display (HMD) is getting popular. However, understanding Quality-of-Experience (QoE) of 360° videos with HMDs is quite challenging because it may be affected by too many factors. This thesis strives to develop models to predict the QoE levels of watching 360° videos with HMDs. To achieve that, first, we implement a 360° video player supporting diverse projection schemes. The 360° video player can support 4K resolution videos at 30 frame-per-second. Then, we conduct a user study to explore the implications of different factors on the QoE of 360° videos. The subjective experiments are designed to exercise: (i) projection schemes, (ii) encoding Quantization Parameters (QPs), and (iii) video genres. We analyze the correlation of the three factors and objective quality metrics with ground-truth QoE levels. Based on the subjective and objective assessments, we construct multiple QoE models to predict the QoE of 360° videos. We also offer our recommendation on model selections under different circumstances.

# Contents


**致謝** **i**

**Acknowledgments** **ii**

**中文摘要** **iii**

**Abstract** **iv**

**1 Introduction** **1**
1.1 Contributions . . . . 2
1.2 Thesis Organization . . . . 3

**2 Background** **4**
2.1 360° Videos . . . . 4
2.2 Virtual Reality . . . . 6
2.3 Viewing Experience . . . . 6

**3 360° Video Player for Diverse Projection Schemes** **9**
3.1 Existing Open-source 360° Video Players . . . . 9
3.2 OSVR-based 360° Video Player . . . . 10
3.3 Diverse Projection Schemes: Design and Implementation . . . . 12
3.4 Performance Measurements . . . . 12

**4 User Study** **15**
4.1 Setup . . . . 15
4.2 Procedure . . . . 18
4.3 Subjective Analysis Results . . . . 19
4.4 Objective Analysis Results . . . . 24

**5 Quality-of-Experience Modeling** **26**
5.1 Potential Factors . . . . 27
5.2 Virtual Reality 360° Video QoE Model . . . . 27
5.3 Validation . . . . 29
5.4 Evaluation . . . . 30

**6 Related Work** **32**
6.1 Players . . . . 32
6.2 User Study . . . . 33
6.3 Modeling . . . . 34

**7 Discussions** **36**
- 7.1 Major Findings . . . 36
- 7.2 Application Scenarios . . . 37
- 7.3 Limitations and Future Directions . . . 37

**8 Conclusion** **39**

**Bibliography** **40**

# List of Figures


1.1 Sample projection schemes: (a) equi-rectangular, (b) adjusted equal-area, and (c) equi-angular cubemap. . . . . . . . . . . . . . . . . . . . . . . . . 1

2.1 Architecture of 360° video streaming system. . . . . . . . . . . . . . . . . 4

3.1 Overview of proposed 360° video player. . . . . . . . . . . . . . . . . . . 10
3.2 Performance comparisons of different projection schemes: (a) 30 and (b) 45 FPS. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

4.1 Sample video frames from: (a) Xmas, (b) Pac-Man, (c) Gorilla, and (d) Chariot Race. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
4.2 Spatial and temporal information of considered sequences. . . . . . . . . 18
4.3 Overview of test procedure for each subject. S and J represent the stimulation and judgment phases. . . . . . . . . . . . . . . . . . . . . . . . . 19
4.4 The outlier samples in Gorilla testing videos. . . . . . . . . . . . . . . . . 20
4.5 MOS scores comparisons between different projection schemes among spatial video genres: (a) simple and (b) complex. . . . . . . . . . . . . . . 21
4.6 Average MOS scores under three QP values. . . . . . . . . . . . . . . . . 21
4.7 MOS scores for each video. . . . . . . . . . . . . . . . . . . . . . . . . . 22
4.8 MOS scores comparison between different temporal genres. . . . . . . . 22
4.9 MOS scores comparison between different spatial genres. . . . . . . . . . 23
4.10 MOS scores comparisons between different temporal genres among spatial genres: (a) simple and (b) complex. . . . . . . . . . . . . . . . . . . . 23
4.11 Objective quality metrics under different QP values: (a) S-PSNR-I and (b) V-PSNR. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
4.12 Quadratic equations for two objective quality metrics: (a) S-PSNR-I and (b) V-PSNR. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24

5.1 Spatial and temporal information of training and testing videos. . . . . . . 30
5.2 New testing 360° videos: (a) Cooking Battle, (b) Hog Rider, and (c) Village. 30

# List of Tables

3.1 Existing Open-source 360° Image/Video Players for HMDs . . . . . . . . 9
3.2 The Average Resource Consumption at 30/45 FPS . . . . . . . . . . . . . 13
3.3 The Time Consumptions of Transformation Mapping and Decoding at 30/45 FPS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

4.1 Descriptions of the Video Sequences . . . . . . . . . . . . . . . . . . . . . 18
4.2 MOS Scores (Standard Deviations) Comparisons . . . . . . . . . . . . . 20

5.1 ANOVA Results for the Potential Factors (Significant Ones are Marked by $*$) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
5.2 Parameters of QoE Models . . . . . . . . . . . . . . . . . . . . . . . . . . 28
5.3 The 3-fold Performance and Significance of Proposed QoE Models . . . . 29
5.4 Descriptions of New Testing Videos . . . . . . . . . . . . . . . . . . . . . 30
5.5 The Performance of Recommended QoE Models . . . . . . . . . . . . . 31

7.1 Comparisons with Other Studies . . . . . . . . . . . . . . . . . . . . . . . 36

# Chapter 1

# Introduction

Recently, Virtual Reality (VR) is getting popular and drawing attention from developers and researchers in both industry and academia. Several online socializing platforms, such as YouTube and Facebook, provide the mature function of sharing 360° videos, which is getting increasingly accepted by people. A market research [15] reports that the market size of VR will reach 40 billion USD in 2020, which shows the potential of VR development. With a wider Field-of-View (FoV) than conventional 2D displays, Head-Mounted Displays (HMDs), such as Oculus Rift, Samsung Gear VR and HTC Vive, allow viewers to look freely in a VR environment and provide a more immersive experience. However, streaming 360° videos requires vast amount network bandwidth, and thus maintaining high Quality-of-Experience (QoE) is critical for retaining viewers.

(a) (b) (c)

Figure 1.1: Sample projection schemes: (a) equi-rectangular, (b) adjusted equal-area, and (c) equi-angular cubemap.

To quantify the QoE levels, human ratings are typically used, where the arithmetic mean of ratings are referred to as the Mean Opinion Score (MOS) in the literature [28]. The subjective quality evaluations are, unfortunately, time-consuming and require a lot of human resources. On the other hand, the objective quality evaluations can be done faster, but previous research shows that the existing objective quality metrics cannot accurately quantify user experience [79]. QoE models are predictive measurements that relate both objective and subjective evaluations. However, there are no existing QoE models in the literature that can measure user experience of watching 360° videos with HMDs.

This thesis strives to provide a QoE model dedicated to watching 360° videos on HMDs. There are some existing QoE models designed for 2D videos. A difference between the QoE models for 360° and 2D videos is the existence of projection schemes. Projection schemes map 360° videos to 2D rectangular ones, which can be compressed by commodity video codecs. Different projection schemes cause diverse shape distortions and pixel densities on 360° videos. Fig. 1.1 presents sample projection schemes: (a) equi-rectangular, which is the most common projection scheme used by 360° video platforms, (b) adjusted equal-area, which reduces the vertical sampling density closed to poles to compensate high horizontal sampling density there, and (c) equi-angular cubemap, which projects to 6 faces of its circumscribed cube.

We carefully develop our QoE models, which requires a 360° video player that supports diverse projection schemes. Unfortunately, there is no existing open-source 360° *video* player supporting multiple projection schemes. Corbillon et al. [9] develop a 360° video player based on Open-Source Virtual Reality (OSVR) [43], which is enhanced by us for heterogeneous projection schemes for the research community [73]. The design of our 360° video player also benefits academia by encouraging easy development of new projection schemes for further research.

To construct a QoE model for 360° video streaming to HMD system, we consider heterogeneous factors as explained subsequently. We recruit 70 subjects to watch 360° videos using Oculus Rift DK2 HMD. We adopt four independent variables in this user study, including: (i) projection schemes, (ii) encoding QP, (iii) spatial video genre, and (iv) temporal video genre. The subjective opinion scores are collected by the Absolute Category Rating (ACR) method for further subjective and objective analysis. According to the results of data analysis, we then propose a QoE model using stepwise linear regression for optimizing user experience.

## 1.1 Contributions

The contributions of this thesis are:

- We realize an open-source 360° video player supporting 4K resolution videos at 30 FPS (frame-per-second) for diverse projection schemes, including equi-rectangular, adjusted equal-area, and equi-angular cubemap. With our proposed 360° video player, we do not need to perform the conversion between projection schemes when presenting 360° videos on HMD. In other words, we skip the conversion that would cause unnecessary video quality reductions for 360° videos. Moreover, our 360° video player allows easy development of new projection schemes.

- We conduct a user study to discuss the impacts of comprehensive factors on QoE of 360° videos, including: (i) projection scheme, (ii) encoding QP, (iii) spatial video genre, and (iv) temporal video genre. In our findings, we recommend the use of equi-angular cubemap projection scheme for 360° video streaming to HMD system since user experiences decrease slightly as QP values increase. In addition, temporal genre has a significant influence on QoE. Slow-paced videos give a better user experience than fast-paced videos.

- We construct QoE linear models based on subjective and objective assessments. We perform a suit of ANOVA tests to find statistical significance on MOS and analyze the correlations between the objective quality metrics for 360° videos and MOS values. Through these processes, we find that the existing objective quality metrics can not be used to predict QoE well. Our proposed QoE model achieves 0.71 Pearson linear correlation coefficient and 0.77 Spearman rank-order correlation coefficient for the testing set while considering projection scheme, encoding QP, and video genres.

## 1.2 Thesis Organization

The rest of this thesis is organized as follows. Chap. 2 provides background on (i) the different entities of 360° video streaming systems, (ii) the applications of 360° videos in virtual reality, and (iii) the state-of-art QoE measurements. This is followed by our proposed 360° video player in Chap. 3, which includes the descriptions of the design and performance. Chap. 4 is about our user study design and the findings of QoE on 360° videos. Chap. 5 includes the derived QoE models and comparisons of its performance. In Chap. 6, we survey the state-of-art related work. We present the major findings of the work, limitations, and its potential in Chap. 7. Finally, in Chap. 8, we give a conclusion.

# Chapter 2

# Background

In this chapter, we introduce the background knowledge of 360° video streaming system, virtual reality, and modern viewing experience measurements.

## 2.1 360° Videos

360° videos, known as omnidirectional videos or spherical videos, allow users to view in all directions. Different from regular 2D videos, 360° videos capture all scenes surrounding cameras. After the scenes are recorded by multiple camera lens, 360° videos have to be stitched into a continuous spherical videos. Such 360° videos burden the transmission costs with high resolutions and bitrates. To alleviate the cost of transmission, we have to project 360° videos onto 2D rectangular ones for encoding. We introduce different entities of 360° video streaming systems as Fig. 2.1, including acquisition, encoding, transmission, and display.

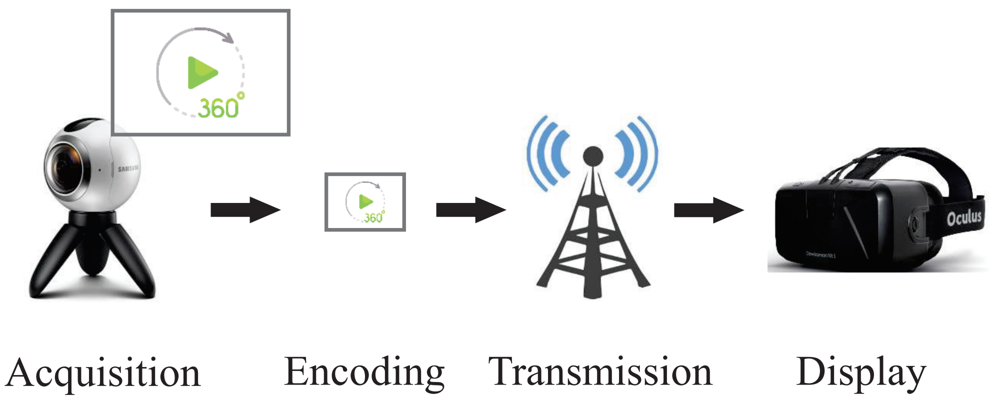



Figure 2.1: Architecture of 360° video streaming system.

**Acquisition.** 360° videos, in order to present all directions of views to users, have to be captured by more than one camera at the same time. By doing so, comprehensive angles are recorded as multiple regular 2D videos aligned with absolute timestamp. These videos need to be stitched as 360° videos. However, unfortunately, the stitching processes involve careful calibration, color correction, and feature mapping, which may degrade qualities of 360° videos by stitching artifacts. Several works [30, 55, 4] in the literature have addressed the stitching issues.

**Encoding.** 360° videos are omnidirectional and spherical. Yet, the existing encoders cannot compress the spherical videos directly. Thus, the efficient projection schemes managing the conversion between 3D and 2D videos is necessary. The process of conversion results in distorted shapes of objects and loss of pixels in original 360° videos. Several studies [53, 38, 14, 72] have proposed different projection schemes for a solution. Nowadays, equi-rectangular is the most popular projection scheme for 360° videos but the shapes of objects near the poles are extremely distorted. To compensate the drawback of equi-rectangular projection scheme, adjusted equal-area one reduces the vertical sampling density to alleviate the shape distortions. Equi-angular cubemap projection scheme is applied by modern YouTube 360° video streaming platform. The projection scheme maps 360° video onto six faces of its circumscribed cube by equal angles of pixel distribution in 3D. We consider the three modern projection schemes into our user study and QoE model.

**Transmission.** Compared to conventional 2D videos, 360° videos require higher resolutions to reach an acceptable viewing experience. It is significantly challenging to transmit 360° videos with large files and high bitrates over bandwidth-restricted networks. Several works have proposed different methods to address transmission issues. For example, Le et al. [37] use MPEG Dynamic Adaptive Streaming over HTTP and spatial representation description to stream tile-based video segments. Graf et al. [19] explore the best transmission strategy based on tile-based adaptive streaming over HTTP. Zare et al. [78] transmit high-resolution tile videos within viewports and other low-resolution tile videos to achieve 30% to 40% bitrate saving. These studies present their streaming systems to maximize efficiently-used network bandwidth.

**Display.** As the technology of 360° camera advances, several online commercial platforms, such as YouTube and Facebook, provide the mature function of sharing 360° videos, which is increasingly accepted and used by people. With a wider FoV and more immersive experience than conventional 2D displays, HMD allows viewers to look freely in a VR environment. Even more, several existing players support HMD to display 360° videos. For instance, Exoplayer [12] and WebVR [70] could be used on Google Cardboard, and OSVR-based video player could be used on Oculus Rift, HTC Vive, and OSVR HDK.

## 2.2 Virtual Reality

**Virtual reality** provides users immersive experience in an artificial virtual environment, which simulates the real world through computer-generated 3D objects. Equipping with a VR headset, users can get audio and visual feedback to completely immerse themselves in VR environments. Within virtual environments, users are able to look around, move around in it, and even interact with virtual 3D objects. Nowadays, 360° images and videos are attractive in virtual reality. A distinctive characteristic of 360° videos is that viewers can be immersed in the recored scenes within videos when looking at every direction of scenes, which is allowed through the existing commercial HMDs. Thus, in this thesis, we focus on QoE of 360° videos in virtual reality.

**Augmented reality** offers users more interactive experiences by 3D virtual objects overlaying onto real components in natural environments. For instance, Pokémon Go [50] is a popular augmented reality mobile game on iOS and Android devices. This game uses the overlaid sensory information of existing environments through a camera and then adds 3D virtual objects to physical world to make a new artificial world. Different from augmented reality, **mixed reality** not only combines 3D virtual objects with the real world, but also allows users to interact with virtual objects as well. For example, Microsoft HoloLens [20] blends digital virtual objects with the real world and then creates an immersive environment. Users can use simple gestures and voice commands to complete different tasks.

## 2.3 Viewing Experience

**Conventional objective quality metrics** are usually designed for measuring qualities of 2D images and videos, such as Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity Index (SSIM), Multi-Scale SSIM (MS-SSIM) and Open Perceptual Video Quality (OPVQ). PSNR measures the pixel-by-pixel coding error between an original content and compressed content, which is unreliable for different contents scenarios [24]. SSIM is used to determine the error of similar characteristics of the contents and outperforms PSNR for perceived image quality approximation [21]. MS-SSIM provides more flexibility in incorporating different resolutions of images compared to SSIM [69]. OPVQ [58] improves the distortion measurement of perceptual evaluation of video quality model standardized by ITU-T Rec. J.247 Annex B [26]. This shows that OPVQ performs significantly better than PSNR. However, these objective quality metrics cannot be used for 360° videos, since the original shapes of objects on 360° videos are distorted. Furthermore, it is challenging for conventional objective quality metrics to quantify 360° videos with different projection schemes.

**Objective quality metrics designed for 360° videos** have been proposed by several studies. For instance, Yu et al. [76] propose to measure the quality between the viewer's FoV of the original video and that of a compressed video according to orientations, which is called Viewport-PSNR (V-PSNR). They also propose a suit of sphere-based PSNR to evaluate the average perceived quality of viewers' by considering the points (i) normally sampled on the sphere space (S-PSNR) or (ii) normally sampled on the sphere space with weights according to the empirical viewing frequency (weighed S-PSNR). Further considering the interpolation problem of S-PSNR metric, JVET evaluates the performance of S-PSNR-I and S-PSNR-NN. They suggest avoiding interpolation in quality metrics whenever possible for future 360° videos coding development. Zakharchenko et al. [77] propose to evaluate the video quality of 360° videos by (i) considering the PSNR on Craster Parabolic Projection (CPP-PSNR), and (ii) modifying the Mean Squared Error (MSE) loss function according to the pixel position in equi-rectangular projection. Sun et al. [62] propose a weighted-to-spherically-uniform PSNR-based (WS-PSNR) method for assessing the video quality of 360° videos. However, these objective qualities [76, 77, 62] does not reflect the user experience.

**Subjective quality metrics** for evaluating QoE levels are typically double stimulus and single stimulus, which are defined by ITU [27, 25]. The double stimulus methods, such as Double Stimulus Continuous Quality Scale and Double Stimulus Comparison Scale, ask subjects to compare the unimpaired reference video sequences and the distorted ones, which take more time and resource to carry out. Compared to the double stimulus methods, the single stimulus methods are more efficient but less consistent in quantifying the individual QoE scores [23]. The reduced consistency of single stimulus method, however, can be eliminated by randomizing the order [49] of testing sequences. Hence, we adopt Absolute Category Rating (ACR), which is one of the single stimulus methods that is suitable for time-independent testing scenarios [27]. However, the subjective quality metrics consume human and material resources.

**QoE for multimedia service** has been studied in several works [3, 1]. QoE is defined as the degree of satisfaction of the users when they use an application or service [29, 51]. QoE is affected by several aspects, including human, system and context influence factors [52], which have been widely studied on traditional 2D videos [80, 61, 22] and 360° videos [63, 79, 36]. In order to design an user study fitting for different scenarios and goals, the pros and cons of within-subjects, between-subjects, and mixed design are discussed [8, 7]. Within-subjects design require each subject to test all conditions of experiments. This design can efficiently isolate the effect of individual differences, but subjects may feel fatigued when being exposed to too many treatments during experiments. In addition, the subjective results may suffer from the order of treatments. Between-subjects design can eliminate the ordering effects and ease subjects' fatigue in experiments, but this design has individual difference problems. Mixed designs combine within-subjects and between-subject design. Researchers can design mixed user study methods according their needs and limitations of different experiments to carefully collect subjective quality data.

# Chapter 3

# 360° Video Player for Diverse Projection Schemes

In this chapter, we are going to introduce existing open-source 360° image/video players for HMDs. Table 3.1 shows a comparison of the players in the literature. We propose our 360° video player testbed to support different projection schemes, which enhances a 360° video player published by Corbillon et al. [9], as well.

## 3.1 Existing Open-source 360° Video Players

Table 3.1: Existing Open-source 360° Image/Video Players for HMDs

| **Player** | **OS** | **HMD** | **Projection Scheme** |
|---|---|---|---|
| | **360° Image Player** | | |
| Upenik et al. [66] | iOS | Portable HMD | Equi-rectangular, Cubemap |
| | **360° Video Player** | | |
| Exoplayer [12] | Android | Portable HMD | Equi-rectangular |
| WebVR [70] | Any OS | Any HMD | Equi-rectangular |
| MP4Client [40] | Any OS | Any HMD | Equi-rectangular |
| Corbillon et al. [9] | Linux, Windows | HDK2 | Equi-rectangular |

Table 3.1 summarizes the representative open-source 360° players. Upenik et al. [66, 67] propose a testbed to conduct a subjective evaluation of 360° images. Subjects wear a wireless HMD equipped with an iPhone 6 to evaluate 360° images represented in equi-rectangular and standard cubemap projection schemes. During the subjective assessment

subjects' head orientations were stored by two coordinates: yaw and pitch. Exoplayer [12] was applied to a HTTP/2-based adaptive 360° video streaming framework by Petrangeli et al. [46, 47, 48]. Exoplayer is a 360° video player for Android, which allows developers to catch head motions from smartphone. WebVR [70] is used to implement tiled-streaming systems by Graf et al. [19] and Ozcinar et al. [45]. WebVR API provides developers access to several HMDs, such as Google Cardboard, Samsung Gear VR, Oculus Rift, and HTC Vive. It allows users to watch 360° videos in a web browser. MP4Client [40] provides developers access to any HMD and OS for watching 360° videos as well. Corbillon et al. [9] implement an OSVR-based 360° video player and release a dataset of head orientations of users watching 360° videos. The OSVR-based player is compatible with Windows OS and any HMD.

Although Upenik et al. [66] proposed a testbed for evaluating 360° images in equi-rectangular and standard cubemap projection schemes, there are not any existing 360° video player supporting different projection schemes. The above-mentioned 360° video players have to convert unsupported projection schemes to equi-rectangular before playing 360° videos. The conversion between projection schemes may lead to unnecessary loss of quality of 360° videos. In order to avoid the conversion and explore the impact of projection schemes on QoE of 360° videos, a 360° video player supporting several projection schemes is needed.

## 3.2 OSVR-based 360° Video Player

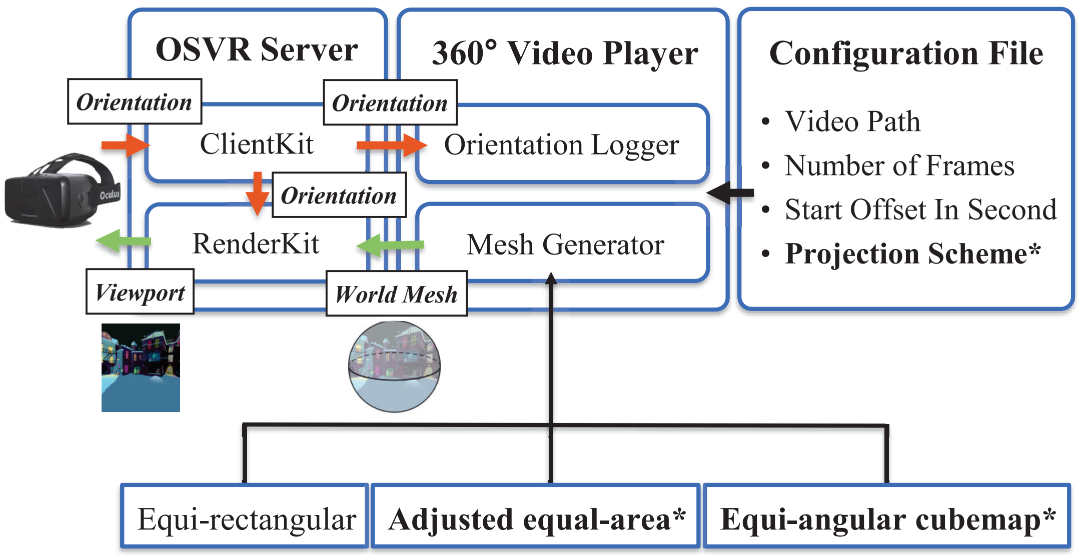



Figure 3.1: Overview of proposed 360° video player.

In order to analyze the behavior of users watching 360° videos, Corbillon et al. [9] implemented an OSVR-based 360° video player and released a dataset of users' navigation patterns. Fig. 3.1 illustrates the architecture of OSVR-based 360° video player, enhanced from the one in Corbillon et al. [9]. Note that our new additions are marked by asterisks.

The OSVR-based 360° video player has two components: (i) 360° video player and (ii) OSVR server. The 360° video player can be configured to play a list of 360° videos by a configuration file. The configuration file facilitates the process of subjective experiments and specifies the path where we store the information of 360° videos and users. The 360° video player has Orientation Logger and Mesh Generator. Orientation Logger gets head orientations of users watching 360° video from HMD and save them to log files. The head orientations are represented as Hamiltons quaternions and aligned with each frame. Mesh Generator firstly produces a sequence of triangles, which are composed of vertex positions in 3D space, to construct a room for rendering. The vertex positions, then, are mapped onto UV plane of 360° videos. The world mesh can be generated and sent to OSVR server.

Open-Source Virtual Reality (OSVR) is an open-source VR platform that aims to enable HMDs from any vendor to be used with any VR software. OSVR provides plugins and APIs to support for several HMDs, such as Razer OSVR HDK2, Oculus Rift, HTC Vive, and FOVE. The OSVR-based 360° video player access HDK2 by ClientKit and RenderKit APIs while OSVR server runs on a local machine. ClientKit is responsible for initializing a client context to use to access HMD. Player can get head orientations from HMD through ClientKit API while users watch 360° videos. RenderKit manages the viewport settings for specified HMD and render the content of viewport onto HMD according to head orientations obtained. With OSVR server, we can easily apply the OSVR-based 360° video player with Oculus Rift we have. Furthermore, in order to reduce latency of rendering viewports, OSVR server also allows us to run the OSVR-based 360° video player on Windows OS with Direct Mode, which directly send viewports to HMD instead of mirroring monitor.

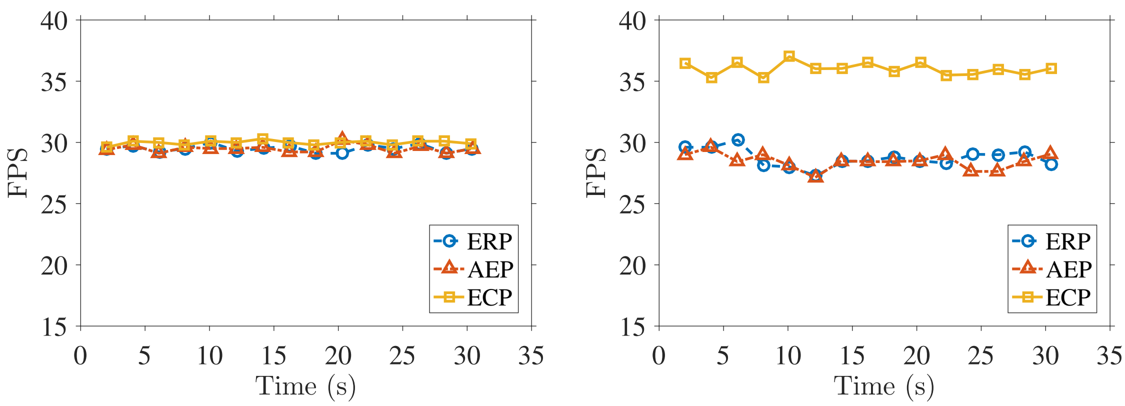
Although the OSVR-based 360° video player allows us to easily conduct subjective experiments with any HMD and OS we prefer, it can not support 360° videos in different projection schemes. The OSVR-based 360° video player only can support equirectangular projection scheme. Without the support of diverse projection schemes, we need to do the conversion of projection schemes before video player starts reading 360° videos in supported projection scheme. This conversion of projection schemes causes the loss of video quality, which results from the interpolation of pixel color values. In order to avoid producing the loss of video quality, we improve the OSVR-based 360° video player to support diverse projection schemes.

## 3.3 Diverse Projection Schemes: Design and Implementation

Fig. 3.1 illustrates the overview of our proposed 360° video player. Different from the OSVR-based 360° video player proposed by Corbillon et al. [9], we improve Mesh Generator of 360° video player to render 360° videos in diverse projection schemes. In the configuration file, we can arrange a list of 360° videos in specified projection schemes for conducting subjective experiments. Our Mesh Generator can map 360° videos onto 3D world space through different projection transformations. With our improved Mesh Generator, we do not need to do the conversion of different projection schemes and thus effectively avoid the unnecessary loss of video quality.

Fig. 3.1 summarizes the implementation of our Mesh Generator. We have two levels of classes to define the relationship between different projection schemes. Superclass, namely Mesh Generator, is responsible for constructing a world space with a sequence of vertex and sending OSVR server the vertex as well as the corresponding UV plane of 360° videos. Subclasses, such as equi-rectangular, adjusted equal-area, and equi-angular cubemap, inherit from Mesh class. The only difference between each projection scheme is projection transformations. Hence, each subclass has own function for producing a sequence of UV 2D points matched with the 3D vertex. The proposed Mesh Generator benefits each projection scheme by reusing the functions of superclass but makes the community easily add more projection schemes based on the architecture of Mesh Generator.

## 3.4 Performance Measurements




(a) (b)

Figure 3.2: Performance comparisons of different projection schemes: (a) 30 and (b) 45 FPS.

Table 3.2: The Average Resource Consumption at 30/45 FPS

| Projection | CPU Load (%) | RAM (MB) | GPU Load (%) | Avg. Output FPS |
|---|---|---|---|---|
| **ERP** | 22.81/23.54 | 1271.82/1271.56 | 15.91/15.47 | 29.49/28.70 |
| **AEP** | 21.80/23.35 | 1267.82/1271.58 | 16.66/15.81 | 29.49/28.45 |
| **ECP** | 20.38/22.56 | 967.65/971.57 | 12.72/14.81 | 29.98/36.01 |

Table 3.3: The Time Consumptions of Transformation Mapping and Decoding at 30/45 FPS

| Projection | Transformation Mapping (ms) | Avg. Decoding Time per Frame (ms) |
|---|---|---|
| **ERP** | 4/4 | 7.45/7.06 |
| **AEP** | 4/4 | 7.90/6.44 |
| **ECP** | 2/2 | 6.44/5.03 |

We report the performance of our 360° video player with a sample 4K video, Chariot Race; other videos lead to similar results. The following metrics are considered:

- *CPU Load.* The CPU utilization in percentage.
- *RAM usage.* The physical memory usage.
- *GPU Load.* The GPU utilization in percentage.
- *FPS.* The number of rendered frames per second.

The video is converted into 3 projection schemes: Equi-Rectangular Projection (ERP), Adjusted Equal-area Projection (AEP), and Equi-angular Cubemap Projection (ECP). 4K resolutions of ERP, AEP, and ECP videos are 3840x1920, 3840x1920, and 2880x1920, respectively. We encode the videos with FFmpeg [13] into H.264 files at QP value of 30. To investigate the maximum FPS the 360° video player achieves, we also prepare the 360° videos at 30 and 45 FPS. The results are reported in Table 3.2.

Our 360° video player can support 4K resolution videos at 30 FPS. Fig. 3.2 presents the FPS achieved by different projection schemes. Fig. 3.2(a) shows that the rendered frame rates for all projection schemes can reach about 30 FPS throughout the entire playback time. In Fig. 3.2(b), we find that the rendered frame rates for ECP can reach around 36 FPS for the video at 45 FPS. This is because the algorithm of projection scheme conversion for ECP involves simpler math operations, compared to ERP and AEP. In addition, ECP has less number of pixels for 4K resolution video. Therefore, ECP has lower computation cost. For ECP, the RAM usages and the GPU utilizations are about 965 MB and 14% respectively.

In addition, we measure the time consumption of projection transformation mapping and decoding video frames. The detailed comparisons are listed in Table 3.3. In order to render 360° videos onto a HMD, we have to map the UV 2D points to 3D vertex on the sphere. ERP and AEP consume 4 ms to perform transformation mapping while ECP consumes 2 ms. The average decoding time per frame for each projection scheme is reported as well. The results indicate that our additional implementation for new projection schemes do not degrade the performance of 360° video player.

# Chapter 4

# User Study

In this chapter, we will present the testbed in our subjective experiment and introduce the variables we select to explore the impacts on QoE of 360° videos from three aspects, namely streaming system design, video codec, and content genres. We are also going to describe the test video sequences and the recruited subjects. Each subject follows the test procedure to evaluate the quality of 360° videos.

## 4.1 Setup

In this section, we introduce our testbed for both subjective and objective assessment of 360° videos on HMD. The main purpose of our 360° video testbed is to present the test video sequences we prepared and to collect subjective opinion scores. In this way, we also store the viewing orientations of the subjects when they watch 360° videos on HMD. This testbed allows us to assess the visual quality of 360° videos using the objective quality metrics, such as S-PSNR-I and V-PSNR [76], based on the recorded orientations. We describe the details of the three components of our testbed as follows, including: (i) HMD, (ii) 360° video player, (iii) OSVR server.

- **HMD.** We use Oculus Rift DK2 [41] as the HMD worn by the subjects for watching 360° videos. The HMD is set up on a PC with Intel i7-3770 CPU, NVIDIA GTX 1060 GPU, 16 GB RAM and Windows 10 OS.

- **360° video player.** We implement a 360° video player developed in C++ language, which is used to display test video sequences to HMD. In addition to supporting diverse projection schemes, the player is connected with OSVR server to store the subjects' head motion data aligned with each frame, as well.

- **OSVR server.** We set up an OSVR server installed with OSVR Oculus plugin [44] to access the tracking sensor on Oculus Rift DK2 HMD. The player can get the last viewer's orientations through OSVR server API and further render the viewport content to HMD. In the meantime, the viewer's head motion data is logged into text files.

While the subjects watch the 360° videos, we record the subjects' head motion data (yaw and pitch) aligned with video timestamps at frame level. It should be noted that the test video sequences are placed into a local hard disk to ensure smooth playback. That is, we do not consider the impact of transmission bandwidth on QoE.

For the objective assessment, we adopt S-PSNR-I and V-PSNR [76] as objective quality metrics to evaluate the perceived quality of 360° videos. We employ 360Lib tool [34] to compute the objective quality of 360° videos based on the collected subjects' orientations.

There are several aspects influencing QoE of 360° videos on HMD. We consider the three aspects, including streaming system design, video codec parameters, and video genres. We describe the variables we select among each aspect in detail.

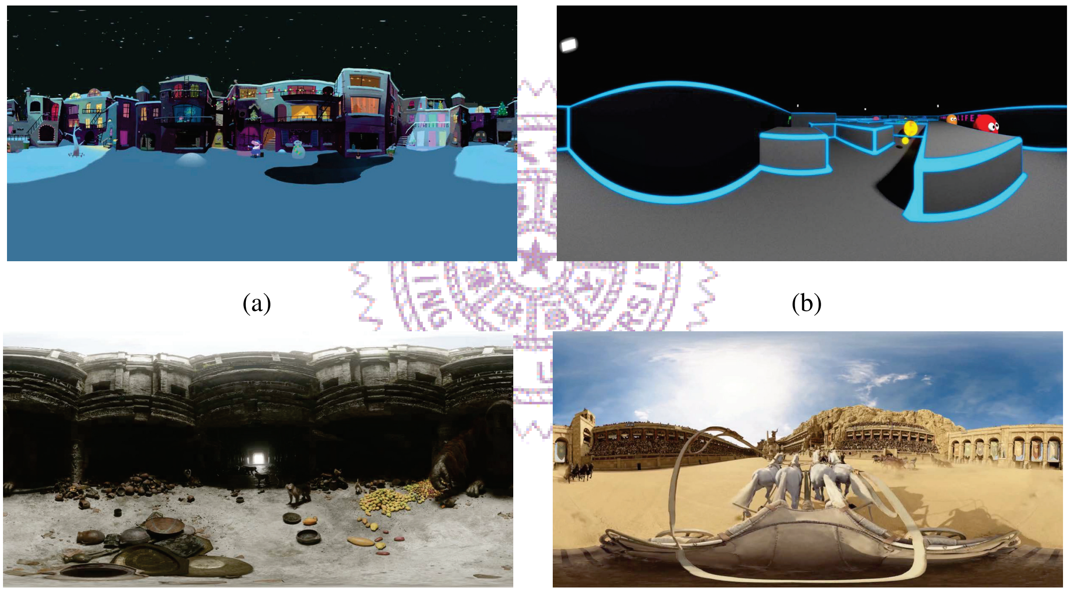
- **Projection scheme.** In order to provide a good user experience with 360° videos, streaming system need to consume vast resource to send 360° videos with high quality. Several solutions to maximize efficiently-used network bandwidth are proposed, such as viewport adaption [10], tiling [11], and different projection usages [16]. We aim to explore the optimal projection scheme for streaming system providing optimal QoE of 360° videos on HMD. We give attention to three types of projection schemes: (i) equi-rectangular, which is the most commonly-used projection scheme, (ii) adjusted equal-area, which properly adjusts the line re-sampling rate along the poles to improve compression efficiency [35], and (iii) equi-angular cubemap, which has better quality at equator as opposed to standard cubemap [75].

- **Encoding quantization parameter.** Given a limited network capacity, video codec has several parameters to restrict file size for transmitting smoothly, such as resolution, bitrate, and QP. Adjusting QP values is one way to achieve target bitrates. However, as QP value rises, the loss of video quality is increased. We focus on the impacts of encoding QP on QoE of 360° videos and how the interaction effects among QP, projection scheme, and content genre influence user experience. We consider the QP values of 22, 30, and 38, respectively corresponding to low, medium, and high video quality.

- **Spatial and temporal video genres.** Different spatial and temporal complexities of 360° videos affect the compression efficiency. In addition, 360° videos can be represented in diverse projection schemes, which result in different degrees of distortion. Some studies [56, 57] indicate QoE of 360° videos is affected by video genres.

However, there is no comprehensive research studying the impacts of temporal and spatial genres of 360° videos on QoE. We have four 360° videos categorized into two groups of simple and complex spatial genres. Among each group, we further group them by slow-paced and fast-paced temporal genres.

In our user study design, we have four independent variables: (i) projection scheme, (ii) encoding QP, (iii) spatial video genre, and (iv) temporal video genre. The dependent variable of user study is MOS collected from 60 subjects watching 360° videos on HMD. To alleviate subjects' fatigue and simplify the interaction effects among the independent variables, we adopt mixed-design [6] for our user study. Projection scheme, encoding QP, and temporal genre serve as within-subjects factors while spatial genre serves as a between-subjects factor. Hence, each subject only evaluates one of spatial groups categorized by simple and complex spatial genres.



(a) (b) (c) (d)

Figure 4.1: Sample video frames from: (a) Xmas, (b) Pac-Man, (c) Gorilla, and (d) Chariot Race.

**Sequences.** We download four 360° videos from YouTube with a resolution of 3840x1920 at 30 fps. These videos are in ERP, and Fig. 4.1 shows sample frames. In order to alleviate subjects' fatigues, we extract 30 seconds from each of them as the references used in our experiment. Table 4.1 summarizes the descriptions of the video sequences. Spatial and Temporal genres are used to categorize the four videos. Fig. 4.2 shows the SI and TI values computed on the luminance plane of 360° videos [27], which demonstrates that the videos are reasonably classified into different content genres. Each reference sequence is converted into 3 different projection schemes, namely ERP, AEP, and EAC, by an open-source 360Lib tool [31]. We then encode the reference videos with H.264 encoder using

Table 4.1: Descriptions of the Video Sequences

| **Video** | **Video Genre** | **Used Segment** | **YouTube Video Id** |
|---|---|---|---|
| Xmas | Simple, slow-paced | 0:10 - 0:40 | XiDRZfeL_hc |
| Pac-Man | Simple, fast-paced | 0:10 - 0:40 | p9h3ZqJa1iA |
| Gorilla | Complex, slow-paced | 0:08 - 0:38 | dKj4PDldebc-U |
| Chariot Race | Complex, fast-paced | 0:02 - 0:32 | jMyDqZe0z7M |

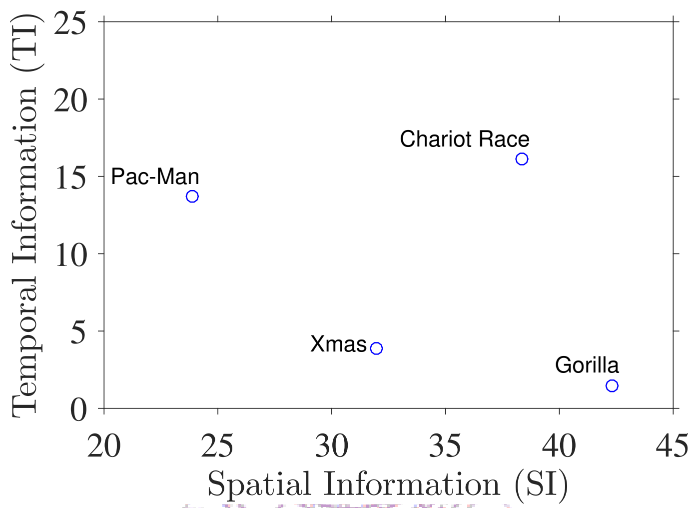



Figure 4.2: Spatial and temporal information of considered sequences.

FFmpeg [13] in three different QP values of 22, 30, and 38. Finally, there are 36 compressed video sequences in total for objective and subjective assessment. While converting the reference sequences into different projection schemes, we make adjustments to the resolution for each projection scheme to ensure a fair comparison as much as possible. We approximate the number of pixels for each projection scheme as the recommendation [33]. We set a resolution with 2880x1920 for EAC and a resolution with 3328x1664 for ERP and AEP. Note that only ERP videos are supported by the current commercial 360° cameras. Even the ECP videos downloaded from YouTube are converted from ERP.

## 4.2 Procedure

**Subjects.** We recruit 60 subjects in our user study. The subjects, 26 females and 34 males, are aged between 19 to 36. There are 23 subjects who have not had any experience with watching 360° videos in a VR environment. Before the subjective assessment, all subjects are asked to complete a testing task to confirm that they do not have visual impairment which may affect their ability to compare the quality of video sequences.

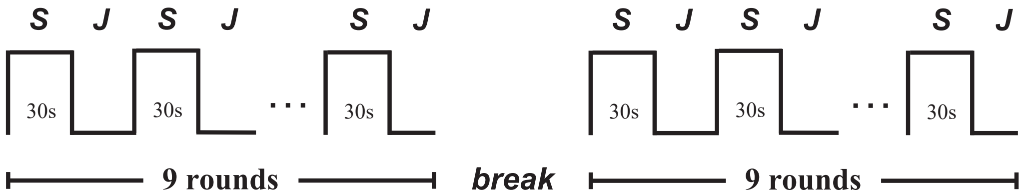



Figure 4.3: Overview of test procedure for each subject. S and J represent the stimulation and judgment phases.

First, we give a training session to the subjects to familiarize them with watching 360° videos on HMDs. The video used in the training session is different from the ones of testing session. We adopt ACR method shown as Fig. 4.3 to collect subjective opinion scores. Before the subjective assessment, we divide the test sequences into two groups by spatial genre and then randomize the presenting order of the test sequences for each subject. Subjects are randomly arranged into two spatial groups and score 18 test sequences. In order to alleviate subjects' fatigues, subjects have a 1-minute break after completing 9 rounds of evaluations. In each round, the subject is asked to watch a 360° video in stimulation phase and then remove the HMD to fill an questionnaire to record the overall experience about the 360° video in judgment phase. To avoid confusing subjects and reducing the user study duration, we only ask a single question: How is your overall experience about this 360° video? In our ACR method, we use a 9-point scale ranging from 1 to 9 to give a more precise rating than a 5-point scale [27]. Note that 9 points represent that subjects have the best overall user experience with the video. In addition, the subjects have unlimited time to answer the several rating questions. After finishing the subjective assessment, we collect 1080 subjective opinion scores from 60 subjects. We perform the objective assessment based on the recored orientations to evaluate the objective quality of 360° videos by S-PSNR-I and V-PSNR metrics.

## 4.3 Subjective Analysis Results

A total of 1080 subjective opinion scores were recorded in our experiments (18 sequences $\times$ 60 subjects). We consider the individual scores outside of 1.5 times of interquartile range as outliers [39, 61]; for example, the outlier samples are marked by red plus in Fig. 4.4. The Gorilla videos from 1 to 9 are the combinations of projection schemes and encoding QPs. After all, 42 outlier samples are removed. The distribution of MOS scores is between 4.5 and 8. Note that all the corresponding 95% confidence intervals are given as errorbars whenever applicable.

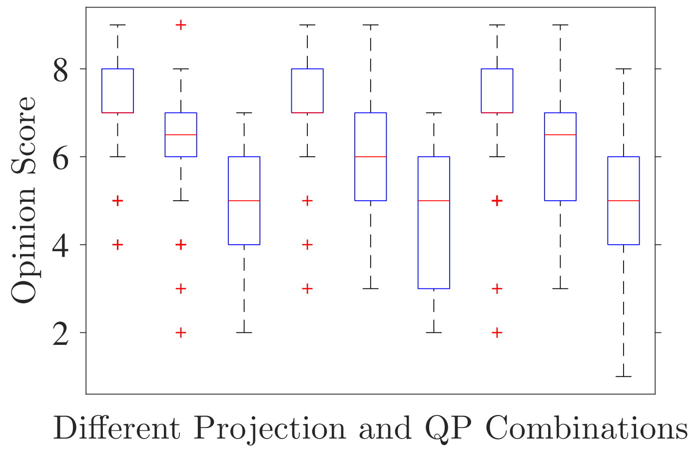


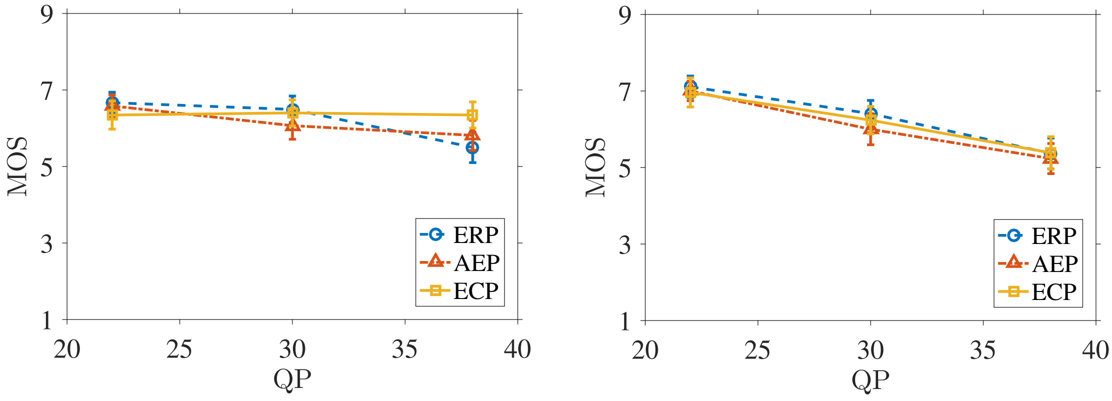

Figure 4.4: The outlier samples in Gorilla testing videos.

Table 4.2: MOS Scores (Standard Deviations) Comparisons

| Projection | Encoding QP | | |
|---|---|---|---|
| | 22 | 30 | 38 |
| ERP | **6.884 (0.204)** | **6.451 (0.250)** | 5.433 (0.289) |
| AEP | 6.789 (0.200) | 6.033 (0.273) | 5.521 (0.288) |
| ECP | 6.640 (0.276) | 6.319 (0.253) | **5.844 (0.293)** |

**Projection schemes alone have no impact on QoE of 360° videos.** Table. 4.2 presents the comparisons of MOS scores among different projection schemes and encoding QPs. Within each encoding QP, projection scheme alone do not incur significant impacts on QoE. However, we find that the interaction effect between projection scheme and encoding QP is significant on QoE of 360° videos, which $p$-value is 0.0008. ECP provides a better experience for subjects watching 360° videos at QP value of 38. This can be attributed to the fact that ECP has less distortions near the poles than ERP and AEP. Instead, ECP allocates uniform pixels for the whole sphere and lower the shape distortions of objects on 360° videos. Thus, MOS scores of ECP decrease slower as QP value increases.

**ECP gives the best QoE for 360° videos with simple spatial video genre.** Fig. 4.5 shows the MOS scores of simple and complex spatial video genres of 360° videos. We observe that the performance of ECP provides the best user experience for the 360° videos with simple video genre in Fig. 4.5(a). This is because simple videos are easy to be compressed and ECP reduces the shape distortions of objects. For example, at the controlled QP value of 38, the average file size of Pac-Man videos is 4.285 MB while that of Char-




(a)

(b)

Figure 4.5: MOS scores comparisons between different projection schemes among spatial video genres: (a) simple and (b) complex.

iot Race videos is 16.270 MB. Thus, simple ECP videos provide the better QoE of 360° videos when the video qualities are low. All project schemes work for 360° videos with complex video genre in Fig. 4.5(b).

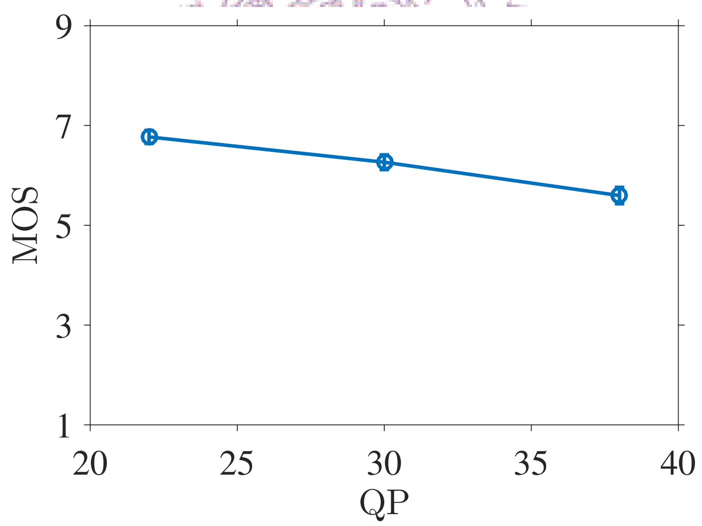



Figure 4.6: Average MOS scores under three QP values.

**QoE of 360° videos decreases as encoding QP rises.** Fig. 4.6 plots the average MOS scores under different QP values for all 360° videos. It demonstrates a negative correlation between the QP value and the MOS score, which results from the video quality is affected by encoding QP. When the encoding QP is increased, the details of videos is discarded so that 360° videos get low qualities. This results correspond with the observations of the studies [79, 63, 65, 64, 57, 56].

**Encoding QP has different effects on different videos.** Fig. 4.7 presents the MOS scores for 360° videos. We perceive that each 360° video has its own pattern of MOS

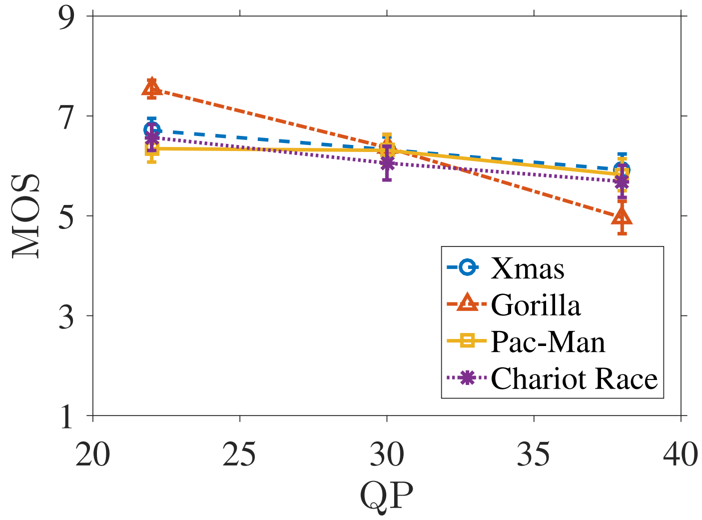



Figure 4.7: MOS scores for each video.

scores under different QP values. For example, the Gorilla video clearly has a steeper pattern than others. In the Gorilla video, gorillas surround the viewers. A characteristic of the Gorilla video is its slow-paced movement but complex spatial scene. Such video genres have the lowest MOS scores at 38 QP value while having the highest MOS scores at 22 QP value. The results indicate that the interaction effect between encoding QP and video genres influences on user experience.

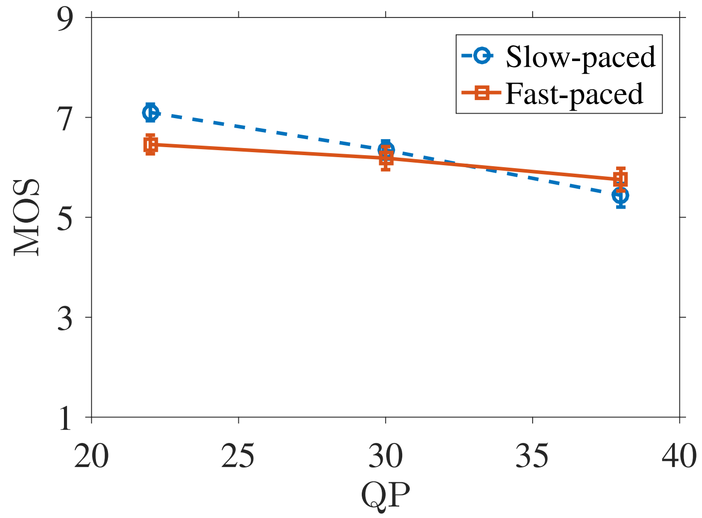



Figure 4.8: MOS scores comparison between different temporal genres.

**Slow-paced videos have more influence on QoE than fast-paced videos.** Fig. 4.8 compares the MOS scores between slow-paced and fast-paced 360° videos with different qualities. It illustrates that subjects have clearly different experiences while watching slow-paced videos of different qualities. Moreover, fast-paced videos do not have the contrast of QoE between different qualities. This is because user experience is sensitive

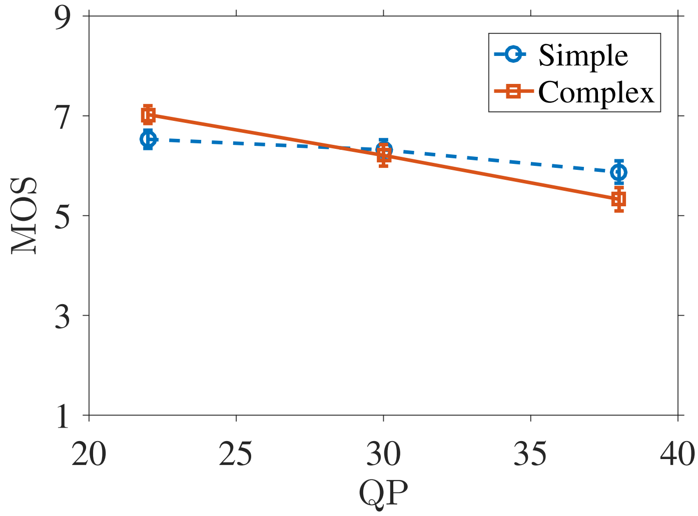



Figure 4.9: MOS scores comparison between different spatial genres.

to the high and low qualities of slow-paced videos instead of fast-paced videos.

**Complex spatial genre has more impact on QoE than simple one.** Fig. 4.9 shows the comparison of MOS scores between simple and complex spatial genres, which demonstrates that complex spatial genre videos produce a great influence on QoE when encoded at different qualities. In contrast with simple spatial scene, the spatial details of complex scene are difficultly saved when QP values are high. Hence, compared to simple spatial genre, complex one provides lower MOS scores at 38 QP value but higher MOS scores at 22 QP value.

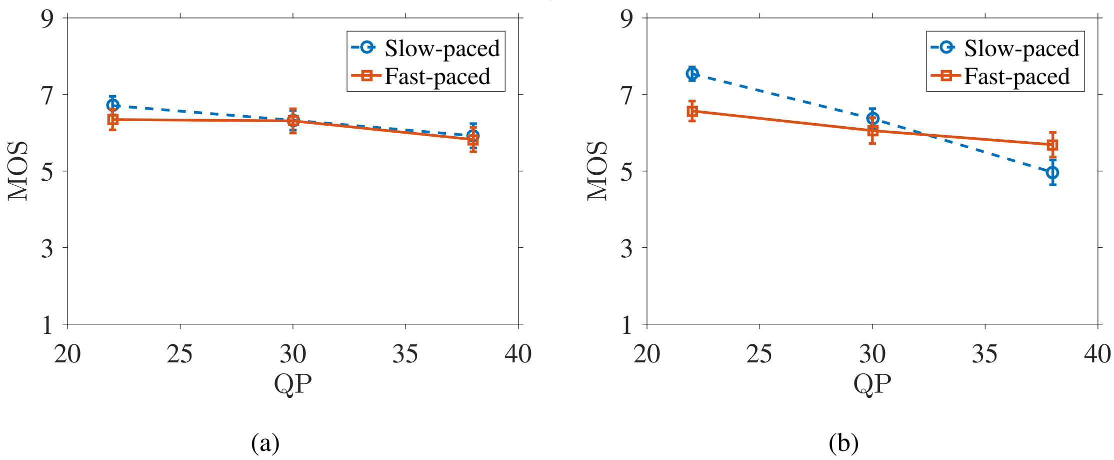



Figure 4.10: MOS scores comparisons between different temporal genres among spatial genres: (a) simple and (b) complex.

**Temporal video genres only have impact on QoE of the complex videos.** Next, we take a closer look at the interaction effects among encoding QP, spatial genre, and temporal genre. Fig. 4.10 plots the MOS scores between different temporal genres in each spatial genre group. It shows that temporal genres only affect QoE of the complex spatial videos, except for the simple spatial videos. This is because the visual qualities of complex spatial scenes are easily distorted and impaired at high QP values.

## 4.4 Objective Analysis Results

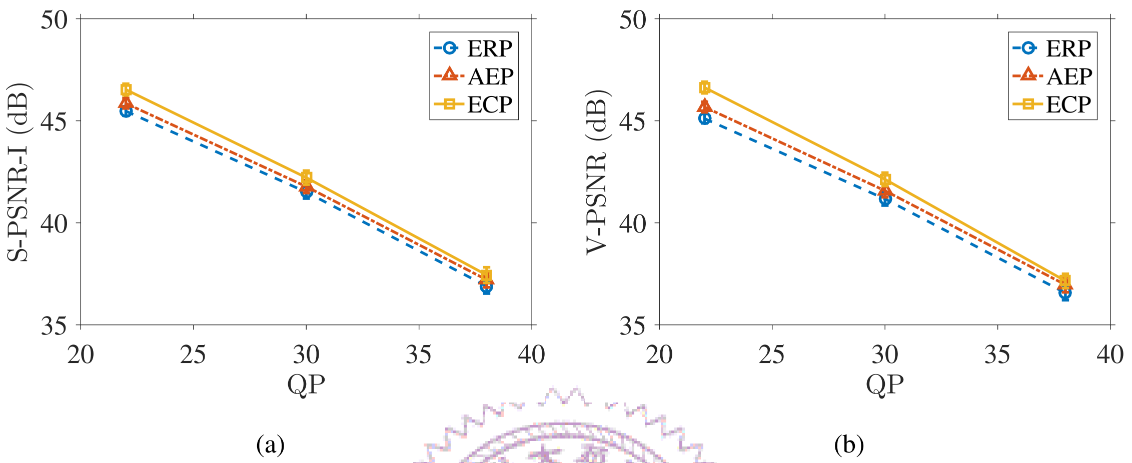



Figure 4.11: Objective quality metrics under different QP values: (a) S-PSNR-I and (b) V-PSNR.

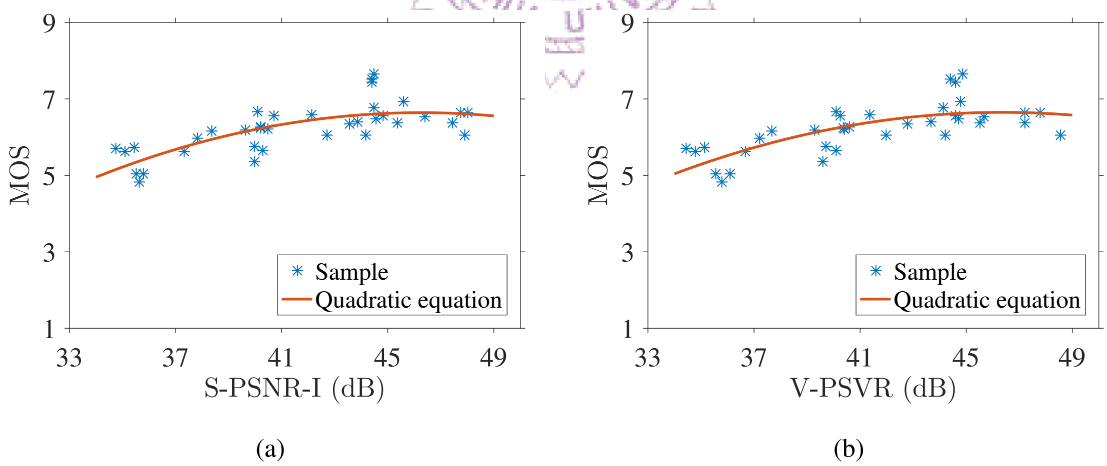



Figure 4.12: Quadratic equations for two objective quality metrics: (a) S-PSNR-I and (b) V-PSNR.

**Objective quality metrics can not be used to predict QoE.** Fig. 4.11 depicts the average objective metric values under three QP values for 4 videos. This figure shows that the average objective quality metrics, including S-PSNR-I and V-PSNR, are highly consistent and have similar patterns at different QP values for all 360° videos. We also observe that EAC significantly outperform the other projection schemes, especially at high quality videos, for both objective quality metrics. The results are consistent with the findings from our subjective analysis. This is because EAC eliminates the great distortions at the poles, compared to ERP and AEP projection schemes. However, as we plot the MOS scores over different objective quality metrics and perform quadratic regressions for modeling QoE in Fig. 4.12, the results of quadratic-polynomial fit are not good to approximate MOS scores. R-square for S-PSNR-I and V-PSNR is a mere $0.5512$ and $0.5377$, respectively. Root-mean-square error (RMSE) for S-PSNR-I and V-PSNR is respectively also high to $0.4262$ and $0.4325$. There is no clear correlation found between the MOS scores and each objective quality metric, which indicated that the objective quality metrics are not good indicators for user experience. Thus, a QoE model considering multiple factors is required.

# Chapter 5

# Quality-of-Experience Modeling

In this chapter, we construct QoE models for 360° videos on HMD. We run stepwise regression to choose significant predictive features for the models. The performance of QoE models are compared and reported.

Table 5.1: ANOVA Results for the Potential Factors (Significant Ones are Marked by $*$)

| **Factor** | **DF** | **Sum Square** | **F Ratio** | ***p*-value** |
|---|---|---|---|---|
| $P$ | 2 | 3.4110 | 0.722 | 0.4859 |
| $Q$ | 1 | 151.7627 | 70.784 | $< .0001^*$ |
| $S$ | 1 | 1.7654 | 0.748 | 0.3874 |
| $T$ | 1 | 2.2804 | 0.967 | 0.3259 |
| $P \times Q$ | 2 | 33.2184 | 7.166 | $0.0008^*$ |
| $P \times S$ | 2 | 2.2222 | 0.470 | 0.6250 |
| $P \times T$ | 2 | 0.9990 | 0.211 | 0.8096 |
| $Q \times S$ | 1 | 28.4052 | 12.236 | $0.0005^*$ |
| $Q \times T$ | 1 | 20.3240 | 8.711 | $0.0033^*$ |
| $S \times T$ | 1 | 0.0718 | 0.030 | 0.8616 |
| $P \times Q \times S$ | 2 | 7.5008 | 1.593 | 0.2042 |
| $P \times Q \times T$ | 2 | 11.8743 | 2.528 | 0.0806 |
| $P \times S \times T$ | 2 | 3.0190 | 0.639 | 0.5280 |
| $Q \times S \times T$ | 1 | 15.5842 | 6.660 | $0.0101^*$ |
| $P \times Q \times S \times T$ | 2 | 0.2476 | 0.052 | 0.9490 |
| $q_{spsnr}$ | 1 | 133.9045 | 61.715 | $< .0001^*$ |
| $q_{vpsnr}$ | 1 | 120.6699 | 55.132 | $< .0001^*$ |

## 5.1 Potential Factors

We consider several potential factors through the subjective and objective analysis, including (i) projection scheme, (ii) encoding QP, (iii) video genres, and (iv) objective quality metrics.

In order to determine the factors significantly contributing to the QoE model, we conduct an Analysis of Variance (ANOVA) test. Degrees of freedoms (DFs), sum of squares, F ratios, and $p$-values are listed on Table 5.1. Let $P$ be the projection schemes, $Q$ be the encoding QP values, $S$ be the spatial genres, $T$ be the temporal genres, $q_{spsnr}$ be the S-PSNR-I quality metric, and $q_{vpsnr}$ be the V-PSNR quality metric. In order to determine the factors significantly contributing to the QoE model, we conduct an Analysis of Variance (ANOVA) test for the training dataset. The results in Table 5.1 show that QoE is significantly affected by the 7 factors: (i) $Q$, (ii) $P \times Q$, (iii) $Q \times S$, (iv) $Q \times T$, (v) $Q \times S \times T$, (vi) $q_{spsnr}$, and (vii) $q_{vpsnr}$. All $p$-values are smaller than $0.05$. We take the factors into account to our QoE model when developing our QoE model.

## 5.2 Virtual Reality 360° Video QoE Model

**Algorithm 1** Stepwise Regression Algorithm.

```
Initialize a linear regression model
Examine p-values for each factor
while available factors not included in the model yet have p-values smaller than the
    entering threshold do
    Add factor with smallest p-value into the model
    Re-calculate p-values for individual factors
    if all available factors included in the model have p-values smaller than the stop-
        ping threshold then
        Output the QoE model
    else
        Remove the factor with the worst p-value
        Re-calculate p-values for individual factors
```

In order to construct a robust QoE model, we run stepwise linear regression shown in Algorithm 1. The individual scores are varied. Instead, we regress the MOS scores to predict overall QoE. It adds the most significant factor into the model in each step. We set the enter and leave thresholds as 0.05. Table 5.2 presents the models constructed by different factor sets. In Table 5.2, $\alpha_{i,j}$ are model parameters, where $i$ and $j$ represent the model number and the factor number, respectively. To validate the performance of the

Table 5.2: Parameters of QoE Models

| Model | Parameters |
|---|---|
| ① : $\alpha_{1,1} + \alpha_{1,2}Q$ | $6.22, -3.77$ |
| ② : ① $+ \alpha_{2,3}QS$ | $6.22, -3.77, 1.59$ |
| ③ : ② $+ \alpha_{3,4}QT$ | $6.22, -3.77, 1.59, -1.74$ |
| ④ : ③ $+ \alpha_{4,5}QST$ | $6.22, -3.77, 1.59, -1.74, 1.12$ |
| ⑤ : ④ $+ \alpha_{5,6}PQ$ | $6.22, -3.77, 1.59, -1.74, 1.12, 1.03$ |
| ⑥ : ⑤ $+ \alpha_{6,7}q_{spsnr}$ | $6.22, -3.58, 1.58, -1.73, 1.13, 1.04, 0.34$ |
| ⑦ : ⑤ $+ \alpha_{7,7}q_{vpsnr}$ | $6.22, -3.66, 1.59, -1.73, 1.13, 1.04, 0.19$ |

objective quality metrics on predicting QoE, we add S-PSNR-I and V-PSNR metrics as additional predictive variable into model ⑤. The models adopting the objective quality metrics are called ⑥ and ⑦.

The QoE models can be applied to various streaming systems that: (i) know video genres only, (ii) know video genres and support multiple projection schemes, and (iii) further compute objective quality metrics, including S-PSNR-I and V-PSNR. Models ①, ②, ③, ④ include encoding QP and video genres factors, which can be utilized to the systems that can only retrieve the information of video genres. Model ⑤ supports the systems that can display diverse projection schemes. Models ⑥ and ⑦ can be employed to the systems that have streaming strategies considering objective qualities of 360° videos. We write the seven models below:

$$\text{①} : MOS = \alpha_{1,1} + \alpha_{1,2}Q; \tag{5.1}$$

$$\text{②} : MOS = \alpha_{2,1} + \alpha_{2,2}Q + \alpha_{2,3}QS; \tag{5.2}$$

$$\begin{aligned} \text{③} : MOS = \alpha_{3,1} \; & + \alpha_{3,2}Q \\ & + \alpha_{3,3}QS + \alpha_{3,4}QT; \end{aligned} \tag{5.3}$$

$$\begin{aligned} \text{④} : MOS = \alpha_{4,1} \; & + \alpha_{4,2}Q \\ & + \alpha_{4,3}QS \\ & + \alpha_{4,4}QT + \alpha_{4,5}QST; \end{aligned} \tag{5.4}$$

$$\begin{aligned} \text{⑤} : MOS = \alpha_{5,1} \; & + \alpha_{5,2}Q \\ & + \alpha_{5,3}QS \\ & + \alpha_{5,4}QT \\ & + \alpha_{5,5}QST + \alpha_{5,6}PQ; \end{aligned} \tag{5.5}$$

$$
\text{⑥} : MOS = \alpha_{6,1} + \alpha_{6,2}Q + \alpha_{6,3}QS + \alpha_{6,4}QT + \alpha_{6,5}QST + \alpha_{6,6}PQ + \alpha_{6,7}q_{spsnr}; \tag{5.6}
$$

$$
\text{⑦} : MOS = \alpha_{7,1} + \alpha_{7,2}Q + \alpha_{7,3}QS + \alpha_{7,4}QT + \alpha_{7,5}QST + \alpha_{7,6}PQ + \alpha_{7,7}q_{vpsnr}, \tag{5.7}
$$

where the model parameters can be found in Table 5.2.

## 5.3 Validation

Table 5.3: The 3-fold Performance and Significance of Proposed QoE Models

| Model | Training Set | | | Validation Set | |
|---|---|---|---|---|---|
| | **PLCC** | **SROCC** | ***p*-value** | **PLCC** | **SROCC** |
| ① | 0.7570 | 0.7767 | $< .0001^*$ | 0.6769 | 0.6861 |
| ② | 0.7998 | 0.7959 | $< .0001^*$ | 0.7190 | 0.7158 |
| ③ | 0.8630 | 0.8353 | $< .0001^*$ | 0.7476 | 0.7205 |
| ④ | 0.8877 | 0.8353 | $< .0001^*$ | 0.7723 | 0.7205 |
| ⑤ | 0.9046 | 0.8570 | $< .0001^*$ | **0.7905** | **0.7430** |
| ⑥ | **0.9190** | 0.8901 | $< .0001^*$ | 0.7395 | 0.6414 |
| ⑦ | 0.9159 | **0.8913** | $< .0001^*$ | 0.7455 | 0.6560 |

We employ Pearson linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SROCC) to quantify the correlation between MOS scores and predicted QoE values. We conduct 3-fold cross validation as well. The 3-fold results of regression evaluation are presented in Table 5.3. Models ③, ④, ⑤, ⑥, and ⑦ achieve above 0.86 PLCC and 0.83 SROCC values for the three types of systems. Models ⑥ and ⑦ involving with S-PSNR-I and V-PSNR metrics have the highest PLCC and SROCC for training set. However, model ⑤ has the highest PLCC and SROCC for the testing set. Through a deeper investigation, we find that the difference between each predicted MOS score and each ground-truth is less than 1. **In summary, we recommend model ④ if projection schemes are unknown and model ⑤ otherwise.**

## 5.4 Evaluation

Table 5.4: Descriptions of New Testing Videos

| **Video** | **Video Genre** | **Used Segment** | **YouTube Video Id** |
|---|---|---|---|
| Cooking Battle | Complex, slow-paced | 0:36 - 1:06 | JpAdLz3iDPE |
| Hog Rider | Complex, fast-paced | 0:00 - 0:30 | yVLfEHXQk08 |
| Village | Simple, slow-paced | 0:10 - 0:40 | QXF7uGfopnY |

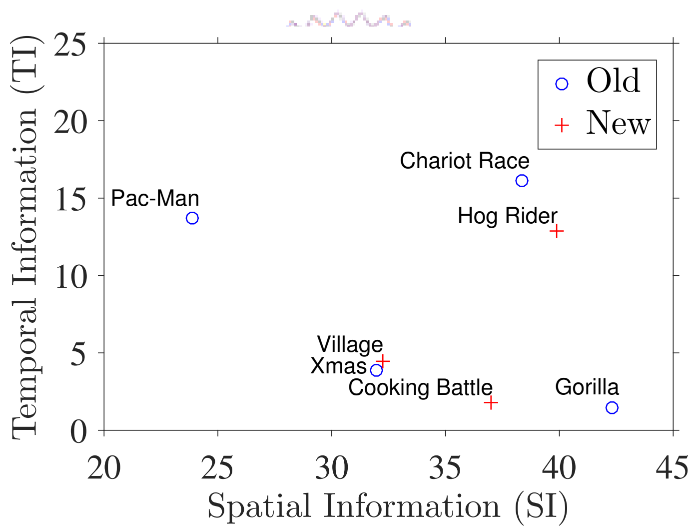



Figure 5.1: Spatial and temporal information of training and testing videos.

(a)


(b)

(c)

Figure 5.2: New testing 360° videos: (a) Cooking Battle, (b) Hog Rider, and (c) Village.

**Setup.** We perform additional evaluations using three new testing videos downloaded from YouTube at 3840x1920 and 30 FPS. Table 5.4 summarizes the descriptions of the testing videos. Fig. 5.1 shows the SI and TI values compared with the old videos. Fig. 5.2

gives sample frames from the new testing videos. We consider three projection schemes and 6 QP values between 24 and 36 for the testing videos, which leads to 18 new testing videos. We recruit 10 additional subjects for the evaluations. The subjects are between 22 to 27 years old; 4 of them are female. All subjects are asked to follow the same test procedure detailed in Sec. 4.2.

Table 5.5: The Performance of Recommended QoE Models

| **Model** | **Testing Set** | |
|---|---|---|
| | **PLCC** | **SROCC** |
| ④ | 0.6880 | 0.7319 |
| ⑤ | 0.7099 | 0.7664 |

**Results.** A total of 180 subjective opinion scores were recorded in the evaluations (18 videos $\times$ 10 subjects). The performance of QoE model ④ and ⑤ are presented in Table 5.5. The PLCC and SROCC scores of model ④ are 0.69 and 0.73. The PLCC and SROCC scores of model ⑤ are 0.71 and 0.77, which are higher than these of model ④. Compared to the PLCC and SROCC scores from the validation set in Table 5.3, the SROCC scores from new videos are slightly higher, while the PLCC scores from new videos are lower. The evaluations on new videos and subjects confirm the robustness of our derived QoE models.

# Chapter 6

# Related Work

In this chapter, we are going to introduce the related works and categorize them into three parts corresponding to our contributions: (i) player, (ii) user study, and (iii) modeling.

## 6.1 Players

**Closed-source 360° video players** are abundantly released in the commerce for watching 360° videos on HMD. For example, Oculus Video [42] and GearVR player [59] are published by Oculus. Oculus Video supports Oculus Rift and Oculus Go HMDs while GearVR player supports wireless Gear VR HMD. Both of the players allow users to watch 4K resolution 360° videos at 30 FPS. However, the video players only support equi-rectangular projection scheme. YouTube player [74] provides immersive VR environments to watch 360° videos as well. Although users can watch 8K resolution 360° videos at 60 FPS on HMD, YouTube stipulates that users only can upload 360° videos in equi-rectangular projection scheme. YouTube then convert the uploaded 360° videos into equi-angular cubemap projection scheme to optimize user experience. Hence, we can only watch equi-angular cubemap projection scheme of 360° videos on HMD. Whirligig [71] is a commercial VR video player, which supports 360° videos in multiple projection schemes, such as equi-rectangular, standard cubemap, and barrel. Nonetheless, Whirligig do not release source code of the player. Thus, we cannot explore the optimal projection scheme for QoE of 360° videos on HMD. We need an open-source 360° video player allowing us easily add additional projection schemes.

**Open-source 360° video players** are used to achieve 360° video streaming in the literature. For instance, Petrangeli et al. [46, 48] apply Exoplayer [12] in a HTTP/2-based adaptive 360° video streaming framework for multiple representation transmission. Exoplayer is a 360° video player for Android, which allows developers to catch head motions from smartphone and supports DASH adaptive playback as well. WebVR player [70] is specified to developed in web browser. Users equipped with Google Cardboard [17] can watch 360° videos in VR environments. GPAC [18] introduces MP4Client [40] video player, which playbacks 360° videos with equi-rectangular projection scheme on HMD. The player can stream tile-based video segments viewed by users from HTTP server to reduce bandwidth consumption. OSVR player [43] provides several plugins and APIs to access any HMD and OS. OSVR releases HDK2 with open hardware as well. However, the open-source video players mentioned above do not support 360° videos with diverse projection schemes. Thus, we enhance the OSVR-based video player, which is proposed by Corbillon et al. [9], to support three modern projection schemes, including equi-rectangular, adjusted equal-area, equi-angular cubemap. Our proposed player benefits the researchers in this domain to easily add new projection scheme for more empirical studies.

## 6.2 User Study

Several works have conducted user study to explore end-side user experience on 360° videos. For instance, Schatz et al. [54] mainly focus on how stalling events affect QoE of 360° videos with HMD and provide some recommendations for upcoming HMD-based VR user study. Tran et al. [64] investigate the influence of encoding QP and video genres on QoE of 360° videos. From their subjective assessments, in order to achieve up to 65% acceptability rate of QoE, the maximum QP values at different resolutions are reported. In addition, the bitrates for good acceptability of QoE are also revealed according to three motion types of 360° videos. Singla et al. [57, 56] categorize 360° videos into three types, namely slow, medium, and fast motions, and also compare the influence of different HMDs on QoE. They reveal that different videos and HMDs have significant impact on QoE. Through the insights of the studies [64, 56, 57], MOS is affected by video genres.

Upenik et al. [66, 67] propose a testbed to perform subjective assessment of 360° images, considering encoding bitrates, projections, and encoders. They analyze the correlation between objective qualities and MOS. Their results show that cubemap results in lower subjective MOS than equi-rectangular at medium encoding bitrates. In addition, they find that the current existing objective quality metrics for 360° images, including S-PSNR, WS-PSNR, and CPP-PSNR, are less correlated with MOS than conventional 2D objective quality metrics. Compared to our work, the previous studies [54, 57, 56, 64, 66, 67] do not explore the optimal projection scheme for 360° videos on HMD and analyze the impacts of video genres, including temporal and spatial characteristics, on QoE of 360° videos.

## 6.3 Modeling

**QoE measurements of conventional 2D videos** are widely studied in the literature. In order to measure QoE, several methods, such as double and single stimulus methods, for evaluating subjective quality are defined by ITU institution [27, 25]. These methods consume a lot of time to recruit human to evaluate video qualities. As opposed to human rating, objective quality metrics, such as Peak Signal-to-Noise Ratio (PSNR) [24] and Structural Similarity Index (SSIM) [5], can measure video quality fast. However, these objective quality metrics do not quantify user experience well [60]. Several studies [68, 2, 61] have proposed QoE models while considering different aspects, such as video genres and codecs. Although the proposed QoE models predict users' perceived quality well, they are still not suitable for evaluating 360° videos. This is because 360° videos have to be projected to 2D rectangular ones for encoding. The transform process causes shape distortions on videos and further influences coding efficiency and user experience. In addition, in order to maintain acceptable QoE, 360° videos require higher resolutions and bitrates than 2D videos, which is limited by network capacity. Hence, we need to propose a model for 360° videos to predict QoE, which helps streaming systems provide better user experience.

**QoE model of 360° videos on HMD** has been recently investigated. Tran et al. [63, 65] evaluate the impacts of the different influence factors, such as encoding QPs and resolutions. It shows that while QoE increases, QP is reduced and resolution is increased. They further investigate the relationship between objective quality and subjective quality of 360° videos, applying the testing procedure for 360° videos defined by JVET [32]. The five objective metrics, namely PSNR, S-PSNR-NN, S-PSNR-I, WS-PSNR, and CPP-PSNR [31, 34], are compared. Their results show that for end-to-end distortion measurement PSNR is the most appropriate metric due to its high correlation with MOS and low complexity. However, Zhang et al. [79] argue that the objective metrics have great room for enhancement of 360° videos due to the observed low values in correlation with MOS scores. They study how different encoders and bitrates affect the quality 360° videos using both subjective and objective metrics. In subjective assessment, they find that x265 encoder outperforms x264 and VP9. In objective assessment, they use four objective quality metrics , namely PSNR, SSIM, VQM, and S-PSNR-I, to evaluate the video quality. Getting both subjective and objective results, they calculate the correlation between the two sets of values. Their result show low correlation between subjective results and objective results, indicating that the four objective quality metrics cannot effectively reflect users' perception on the video quality. Kim et al. [36] propose a VR sickness predictor to address visual-vestibular sensory problem in VR environments. This predictor involving perceptual motion and statistical content features regresses MOS scores and individual sickness scores by support vector machine. Focusing on the impacts of object and camera movements on sickness and MOS scores in VR environments, they generate diversified scenes by Unity 3D engine for their experiments. Compared to the above-mentioned studies, we explore the optimal projection scheme for QoE of 360° videos on HMD and analyze the impacts of video genres, including temporal and spatial characteristics, on QoE. To address the issues, **we develop the first ever QoE model considering diverse projection schemes for watching 360° videos on HMDs.**

# Chapter 7

# Discussions

## 7.1 Major Findings

Table 7.1: Comparisons with Other Studies

| Study | Zhang et al. [79] | Tran et al. [63, 65] | Tran et al. [64] | Singla et al. [57, 56] | This paper |
|---|---|---|---|---|---|
| **Method** | Subjective, Objective | Subjective, Objective | Subjective | Subjective | Subjective, Objective |
| **Encoder** | H.264, H.265, VP9 | H.264 | H.264 | H.265 | H.264 |
| **Encoding Bitrates/QPs** | 0.3-10 Mbps | 22-40 QP | 22-40 QP | 0.5-15 Mbps | 22-38 QP |
| **Encoding Resolutions** | 4K | 720p-4K | 720p-4K | 1080p, 4K | 4K |
| **Projection Schemes** | ERP | ERP | ERP | ERP | ERP, AEP, ECP |
| **Video Genres** | None | | They reveal that different 360° videos affect QoE levels. | | Spatial, Temporal |
| **QoE Model** | Not Developed | | | | The First QoE Models |

We consider multiple factors that may influence the QoE levels of 360° videos with HMDs, including projection schemes, encoding QPs, and video genres. Our findings are consistent with the literature in most aspects. For instance, encoding QP has impacts on QoE of 360° videos [79, 63, 65, 64, 57, 56]: higher QP values result in worse user experience. Another example is that video genres affect the QoE levels of 360° videos [64, 57, 56]. In addition, we list the new findings compared to the literature [79, 63, 65, 64, 57, 56].

- Projection scheme alone has no significant impact on QoE of 360° videos. Yet, equi-angular cubemap projection scheme provides a better experience for subjects watching 360° videos when QP value is higher.

- Slow-paced videos give more diversity of QoE with different QP values than fast-paced videos do. In addition, the QoE of complex 360° videos is more sensitive to QP values.

- QoE model ⑤, which considering projection schemes, video genres, and encoding QP, has the best performance to predict QoE of 360° videos. Moreover, the objective quality metrics do not incur significant impacts on QoE models.

Table 7.1 compares our work with other studies in the literature.

## 7.2 Application Scenarios

Our derived QoE models can be applied in such scenarios.

- **Server storage management.** 360° videos with high resolutions and bitrates require vast amount data storage in a server. Each video, however, has different changes of QoE levels under different QP values. Through our developed QoE models, we can encode videos at maximum QP values for different QoE levels. That is, the data storage in a server can be reduced when we have optimal encoding QP selections for different QoE levels.

- **Acceptable QoE levels delivery.** Maintaining high QoE levels is critical for retaining viewers. When a server streams massive 360° videos of high qualities, it would suffer from limited network bandwidth. Thus, when a server decides how to transmit acceptable QoE levels of 360° videos to viewers, it can adopt the QoE metrics into the transmission mechanism.

Through applying our QoE models that predict QoE levels of 360° videos under diverse QP values, we improve the server data storage managements and deliver acceptable QoE levels of 360° videos.

## 7.3 Limitations and Future Directions

This work can be extended in such directions.

- **Human viewing behavior.** 360° videos allow viewers to watch freely in VR environments. This means that viewers may have various user experiences when watching different parts of 360° videos. In the thesis, we only predict the average QoE levels of 360° videos for all users. We may analyze the human viewing behaviors for individual QoE models.

- **The degree of sickness.** The 360° videos can be classified by the camera setups, which incur different levels of sickness. For instance, although mounting a camera on a roller coaster gives immersive experiences, viewers may suffer from VR sickness while watching the video in HMD. Thus, a VR sickness predictor may be integrated with our QoE models, especially for longer videos.

- **Integration with 360° video streaming systems.** Currently, our QoE models are trained within a controlled experimental environment. However, if we want to integrate our QoE model with real 360° video streaming systems, we have to consider the impact of transmission bandwidth on QoE. The stalling events of watching 360° videos significantly affect user experience [54]. In addition, we can also apply online machine learning for optimizing the performance of QoE models in real time.

By realizing the 360° video player supporting diverse projection schemes and modeling the QoE of 360° video on HMD, we open up new directions for interested researchers to further explore QoE of 360° videos in VR environments.

# Chapter 8

# Conclusion

In this thesis, we implement a 360° video player, conduct a user study, and derive QoE models. Our 360° video player [73] supports 4K resolution 360° videos with diverse projection schemes up to 30 FPS. This allows us and interested researchers to effectively perform user studies and add new projection schemes to the player. We believe this will stimulate more empirical studies in this domain. In contrast to most studies discussing the performance of projection schemes by objective quality metrics, we address the impact of projection schemes on QoE by subjective quality metrics, reveal that equi-angular cubemap projection scheme offers a better QoE of 360° videos at low video qualities, and show the influence of video genres on QoE of 360° videos. User experiences are sensitive to QP values with slow-paced and complex videos. Future streaming systems are recommended to adopt equi-angular cubemap projection scheme and employ different encoding parameters for diverse video genres. Our user study involves 70 subjects and 7 videos in total. Several QoE models are developed using the user study results. We recommend two QoE models for 360° video systems with/without prior knowledge on projection schemes. The QoE model considering projection schemes can achieve 0.71 PLCC and 0.77 SROCC scores for the testing set. We also find that objective quality metrics do not incur significant impacts on QoE models.

# Bibliography


[1] M. Alreshoodi and J. Woods. Survey on QoE/QoS correlation models for multimedia services. *International Journal of Distributed and Parallel Systems*, 4(3), May 2013.

[2] L. Anegekuh, L. Sun, and E. Ifeachor. Encoding and video content based HEVC video quality prediction. *Multimedia Tools and Applications*, 74(11):3715–3738, June 2015.

[3] S. Baraković and L. Skorin-Kapov. Survey and challenges of QoE management issues in wireless networks. *Computer Networks and Communications*, December 2012.

[4] K. Calagari, M. Elgharib, S. Shirmohammadi, and M. Hefeeda. Sports VR content generation from regular camera feeds. In *Proc. of ACM International Conference on Multimedia (MM'17)*, pages 699–707, Mountain View, USA, October 2017.

[5] S. Channappayya, A. Bovik, C. Caramanis, and R. Heath. SSIM-optimal linear image restoration. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 765–768, Las Vegas, USA, March 2008.

[6] G. Charness, U. Gneezy, and M. Kuhn. Experimental methods: Between-subject and within-subject design. *Economic Behavior & Organization*, 81(1):1–8, January 2012.

[7] G. Charness, U. Gneezy, and M. A. Kuhn. Experimental methods: Between-subject and within-subject design. *Economic Behavior & Organization*, 81(1):1–8, January 2012.

[8] L. Christensen, B. Johnson, and L. Turner. Experimental research design. In *Research methods, design, and analysis*, chapter 8, pages 217–248. Allyn & Bacon Boston, MA, 2011.

[9] X. Corbillon, F. De Simone, and G. Simon. 360-degree video head movement dataset. In *Proc. of the 8th ACM on Multimedia Systems Conference (MMSys'17)*, pages 199–204, Taipei, Taiwan, June 2017.

[10] X. Corbillon, G. Simon, A. Devlic, and J. Chakareski. Viewport-adaptive navigable 360-degree video delivery. In *Proc. of IEEE International Conference on Communications (ICC'17)*, pages 1–7, Paris, France, May 2017.

[11] I. D. Curcio, H. Toukomaa, and D. Naik. Bandwidth reduction of omnidirectional viewport-dependent video streaming via subjective quality assessment. In *Proc. of the 2Nd International Workshop on Multimedia Alternate Realities (AltMM'17)*, pages 9–14, Mountain View, California, USA, October 2017.

[12] ExoPlayer : An extensible media player for android, July 2018. https://github.com/google/ExoPlayer.

[13] FFmpeg web page, July 2018. http://www.ffmpeg.org.

[14] C. Fu, L. Wan, T. Wong, and C. Leung. The rhombic dodecahedron map: An efficient scheme for encoding panoramic video. *IEEE Transactions on Multimedia*, 11(4):634–644, June 2009.

[15] Global virtual reality market (hardware and software) and forecast to 2020, June 2018. https://goo.gl/vtFXVK.

[16] Google AR and VR Blog web page, July 2018. https://blog.google/products/google-ar-vr/bringing-pixels-front-and-center-vr-video/.

[17] Google Cardboard, July 2018. https://vr.google.com/cardboard/.

[18] GPAC: Multimedia open source project, October 2018. https://gpac.wp.imt.fr/.

[19] M. Graf, C. Timmerer, and C. Mueller. Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation. In *Proc. of the 8th ACM on Multimedia Systems Conference (MMSys'17)*, pages 261–271, Taipei, Taiwan, June 2017.

[20] Microsoft HoloLens, October 2018. https://www.microsoft.com/en-us/hololens.

[21] A. Hore and D. Ziou. Image quality metrics: PSNR vs. SSIM. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2366–2369. IEEE, 2010.

[22] T. Hoßfeld, L. Skorin-Kapov, Y. Haddad, P. Pocta, V. A. Siris, A. Zgank, and H. Melvin. Can context monitoring improve QoE? a case study of video flash crowds in the internet of services. In *Proc. of IEEE International Symposium on Integrated Network Management (IM'15)*, pages 1274–1277, Ottawa, ON, Canada, May 2015.

[23] C. Hsu, A. Chen, C. Hsu, C. Huang, C. Lei, and K. Chen. Is foveated rendering perceivable in virtual reality? exploring the efficiency and consistency of quality assessment methods. In *Proceedings of the 2017 ACM on Multimedia Conference*. ACM, 2017.

[24] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 44(13):800–801, June 2008.

[25] ITU-R. BT.500-13: Methodology for the subjective assessment of the quality of television pictures. Technical report, January 2012.

[26] ITU-T. J.247 Objective perceptual multimedia video quality measurement in the presence of a full reference. Technical report, August 2008.

[27] ITU-T. P.910 Subjective video quality assessment methods for multimedia applications. Technical report, April 2008.

[28] ITU-T. P.800.1 Mean opinion score (MOS) terminology. Technical report, August 2016.

[29] Vocabulary for performance and quality of service. Standard, ITU Telecommunication Standardization Sector, 2016.

[30] W. Jiang and J. Gu. Video stitching with spatial-temporal content-preserving warping. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*, pages 42–48, Boston, USA, June 2015.

[31] Joint video exploration team, "360Lib", September 2017. `https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/tags/360Lib-2.0.1/`.

[32] AHG8: Suggested testing procedure for 360-degree video. Technical report, Joint Video Exploration Team of ITU Telecommunication Standardization Sector, October 2016.

[33] JVET common test conditions and evaluation procedures for 360° video. Technical report, Joint Video Exploration Team of ITU Telecommunication Standardization Sector, October 2016.

[34] Algorithm descriptions of projection format coversion and video quality metrics in 360Lib. Technical report, Joint Video Exploration Team of ITU Telecommunication Standardization Sector, March 2017.

[35] AHG8: A study on quality impact of line re-sampling rate in EAP. Technical report, Joint Video Exploration Team of ITU Telecommunication Standardization Sector, July 2017.

[36] J. Kim, W. Kim, S. Ahn, J. Kim, and S. Lee. Virtual reality sickness predictor: Analysis of visual-vestibular conflict and VR contents. In *Proc. of IEEE International Conference on Quality of Multimedia Experience (QoMEX'18)*, pages 1–6, Cagliari, Italy, May 2018.

[37] J. Le Feuvre and C. Concolato. Tiled-based adaptive streaming using MPEG-DASH. In *Proc. of the 7th ACM on Multimedia Systems Conference (MMSys'16)*, pages 41:1–41:3, Klagenfurt, Austria, May 2016.

[38] L. Li, Z. Li, X. Ma, H. Yang, and H. Li. Co-projection-plane based 3-D padding for polyhedron projection for 360-degree video. In *Proc. of IEEE International Conference on Multimedia and Expo (ICME'17)*, pages 55–60, Hong Kong, China, July 2017.

[39] R. McGill, J. Tukey, and W. Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, February 1978.

[40] MP4Client video player, October 2018. `https://gpac.wp.imt.fr/player/`.

[41] Oculus Rift web page, July 2018. `https://www.oculus.com/`.

[42] Oculus Video web page, July 2018. `https://www.oculus.com/experiences/rift/926562347437041/`.

[43] Open-Source Virtual Reality web page, July 2017. `https://osvr.github.io/`.

[44] OSVR-Oculus-Rift github, July 2018. `https://github.com/OSVR/OSVR-Oculus-Rift`.

[45] C. Ozcinar, A. De Abreu, and A. Smolic. Viewport-aware adaptive 360° video streaming using tiles for virtual reality. In *Proc. of IEEE International Conference on Image Processing (ICIP'17)*, pages 2174–2178, Beijing, China, September 2017.

[46] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. De Turck. An HTTP/2-based adaptive streaming framework for 360° virtual reality videos. In *Proc. of ACM on Multimedia Conference (MM'17)*, pages 306–314, Mountain View, California, USA, October 2017.

[47] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. De Turck. Improving virtual reality streaming using HTTP/2. In *Proc. of the 8th ACM on Multimedia Systems Conference (MMSys'17)*, pages 225–228, Taipei, Taiwan, June 2017.

[48] S. Petrangeli, J. van der Hooft, T. Wauters, R. Huysegems, P. Alface, T. Bostoen, and F. De Turck. Live streaming of 4K ultra-high definition video over the internet. In *Proc. of the 7th International Conference on Multimedia Systems (MMSys'16)*, pages 27:1–27:4, Klagenfurt, Austria, 2016.

[49] M. H. Pinson and S. Wolf. Comparing subjective video quality testing methodologies. In *Visual Communications and Image Processing 2003*, pages 573–582. International Society for Optics and Photonics, 2003.

[50] Pokémon Go, October 2018. http://www.pokemongo.com/.

[51] Qualinet white paper on definitions of quality of experience, March 2013. https://hal.archives-ouvertes.fr/hal-00977812.

[52] U. Reiter, K. Brunnström, K. De-Moor, M.-C. Larabi, M. Pereira, A. Pinheiro, J. You, and A. Zgank. Factors influencing quality of experience. In S. Möller and A. Raake, editors, *Quality of Experience: Advanced Concepts, Applications and Methods*, chapter 4, pages 55–72. Springer US, March 2014.

[53] J. Sauer, J. Schneider, and M. Wien. Improved motion compensation for 360° video projected to polytopes. In *Proc. of IEEE International Conference on Multimedia and Expo (ICME'17)*, pages 61–66, Hong Kong, China, July 2017.

[54] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo. Towards subjective quality of experience assessment for omnidirectional video streaming. In *Proc. of IEEE International Conference on Quality of Multimedia Experience (QoMEX'17)*, pages 1–6, Erfurt, Germany, May 2017.

[55] R. Silva, B. Feijó, P. Gomes, T. Frensh, and D. Monteiro. Real time 360° video stitching and streaming. In *Proc of ACM SIGGRAPH 2016 Posters (SIGGRAPH '16)*, pages 70:1–70:2, Anaheim, California, July 2016.

[56] A. Singla, S. Fremerey, W. Robitza, P. Lebreton, and A. Raake. Comparison of subjective quality evaluation for HEVC encoded omnidirectional videos at different bit-rates for uhd and fhd resolution. In *Proc. of Thematic Workshops of ACM Multimedia (Thematic Workshops'17)*, pages 511–519, Mountain View, California, USA, October 2017.

[57] A. Singla, S. Fremerey, W. Robitza, and A. Raake. Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays. In *Proc. of IEEE International Conference on Quality of Multimedia Experience (QoMEX'17)*, pages 1–6, Erfurt, Germany, May 2017.

[58] K. Skarseth, H. Bjørlo, P. Halvorsen, M. Riegler, and C. Griwodz. OpenVQ: A video quality assessment toolkit. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 1197–1200. ACM, 2016.

[59] SKYBOX VR Video Player web page, July 2018. `https://www.oculus.com/experiences/gear-vr/1256687334386627/`.

[60] W. Song and D. Tjondronegoro. Acceptability-based QoE models for mobile video. *IEEE Transactions on Multimedia*, 16(3):738–750, April 2014.

[61] W. Song, Y. Xiao, D. Tjondronegoro, and A. Liotta. QoE modelling for VP9 and H.265 videos on mobile devices. In *Proc. of ACM international conference on Multimedia (MM'15)*, pages 501–510, Brisbane, Australia, October 2015.

[62] Y. Sun, A. Lu, and L. Yu. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. *IEEE Signal Processing Letters*, 24(9):1408–1412, Sept 2017.

[63] H. Tran, N. Ngoc, C. Bui, M. Pham, and T. Thang. An evaluation of quality metrics for 360 videos. In *Proc. of IEEE 9th International Conference on Ubiquitous and Future Networks (ICUFN'17)*, pages 7–11, Milan, Italy, July 2017.

[64] H. Tran, N. Ngoc, C. Pham, Y. Jung, and T. Thang. A subjective study on QoE of 360 video for VR communication. In *Proc. of IEEE 19th International Workshop on Multimedia Signal Processing (MMSP'17)*, pages 1–6, Luton, UK, October 2017.

[65] H. Tran, C. Pham, N. Ngoc, A. Pham, and T. Thang. A study on quality metrics for 360 video communications. *IEICE TRANSACTIONS on Information and Systems*, 101(1):28–36, January 2018.

[66] E. Upenik, M. Řeřábek, and T. Ebrahimi. Testbed for subjective evaluation of omnidirectional visual content. In *Proc. of Picture Coding Symposium (PCS'16)*, pages 1–5, Nuremberg, Germany, December 2016.

[67] E. Upenik, M. Rerabek, and T. Ebrahimi. On the performance of objective metrics for omnidirectional visual content. In *Proc. of IEEE International Conference on Quality of Multimedia Experience (QoMEX'17)*, pages 1–6, Erfurt, Germany, May 2017.

[68] Y. Wang, K. Zhu, J. Wu, and Y. Zhu. Content aware video quality prediction model for HEVC encoded bitstream. *Multimedia Tools and Applications*, 76(18):19191–19209, March 2017.

[69] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 2, pages 1398–1402. IEEE, 2003.

[70] WebVR: Bringing virtual reality to the web, July 2018. `https://webvr.info/`.

[71] Whirligig Player web page, July 2018. `http://www.whirligig.xyz/`.

[72] M. Xiao, S. Wang, C. Zhou, L. Liu, Z. Li, Y. Liu, and S. Chen. Miniview layout for bandwidth-efficient 360-degree video. In *Proc. of ACM International Conference on Multimedia (MM'18)*, pages 914–922, Seoul, Republic of Korea, October 2018.

[73] S. Yao. A 360 degree video player and head movement logger for an HMD using the OSVR API, August 2018. `https://github.com/shunhuaiyao/360Degree_Head_Movement_Dataset/tree/win32/`.

[74] YouTube, August 2018. `https://www.youtube.com/`.

[75] YouTube Engineering and Developers Blog web page, July 2018. `https://youtube-eng.googleblog.com/2017/03/improving-vr-videos.html`.

[76] M. Yu, H. Lakshman, and B. Girod. A framework to evaluate omnidirectional video coding schemes. In *Proc. of IEEE International Symposium on Mixed and Augmented Reality (ISMAR'15)*, pages 31–36, Fukuoka, Japan, September 2015.

[77] V. Zakharchenko, K. P. Choi, and J. H. Park. Quality metric for spherical panoramic video. In *Proc. of SPIE Optical Engineering+ Applications*, pages 99700C–99700C, San Diego, California, United States, September 2016.

[78] A. Zare, A. Aminlou, M. Hannuksela, and M. Gabbouj. HEVC-compliant tile-based streaming of panoramic video for virtual reality applications. In *Proc. of ACM International Conference on Multimedia (MM'16)*, pages 601–605, Amsterdam, The Netherlands, October 2016.

[79] B. Zhang, J. Zhao, S. Yang, Y. Zhang, J. Wang, and Z. Fei. Subjective and objective quality assessment of panoramic videos in virtual reality environments. In *Proc. of IEEE International Conference on Multimedia and Expo Workshops (ICMEW'17)*, pages 163–168, Hong Kong, China, July 2017.

[80] Y. Zhu, A. Hanjalic, and J. A. Redi. QoE prediction for enriched assessment of individual video viewing experience. In *Proc. of ACM international conference on Multimedia (MM'16)*, pages 801–810, Amsterdam, The Netherlands, October 2016.