

Rate-Distortion Modeling of Synthesizer Immersive Video for 6DoF Interaction

Yuan Jun Sun

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

email address or ORCID

Abstract—This document is a model and instructions for \LaTeX . This and the `IEEEtran.cls` file define the components of your paper [title, text, heads, etc.]. ***CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

Recently, Virtual Reality (VR) becomes increasingly more popular. It enables a wide array of novel applications in many domains, such as video streaming, computer games, occupational training, healthcare, manufacturing, etc. The market research also reports that foresee explosive growth of the VR market in the upcoming years [1]. More and more companies devote their effort to the VR industry such as Meta [2].

One way to classify the VR applications is through the different *interaction techniques*, including *3DoF* and *6DoF* interactions. Because of the high popularity of 360° video streaming services, most users are familiar with the *3DoF* (Degree-of-Freedom) interactions, in which a user's viewport is determined by his/her head/HMD *orientation*. In the *3DoF* VR application, the user's position in his/her coordinate such as standing up and walking around, his/her HMD viewport would not reflect the changes of positions. Therefore, the user will *not* feel he/she is moving in the virtual world, leading to an inferior *immersive* user experience. In *6DoF* interactions, the application will render the viewports depending on the user position and orientation. Different from the *3DoF* interaction, the *6DoF* interaction allows users to walk around the virtual world which optimizes the immersive user experience. Fig. 1 illustrates the difference between *3DoF* and *6DoF* interactions.

Supporting *6DoF* Extended Reality (XR) using 360° videos is not an easy task, because, for every single position, a new 360° video needs to be captured. Even if we deploy dense 360° cameras, users may still miss smooth transitions at the positions between any two adjacent cameras. Hence, more descriptive 3D representations are required for enabling the truly immersive experience of *6DoF* VR applications. Recently, MPEG-I (Moving Picture Expert Group–Immersive Group) has been actively developing MPEG Immersive Video (MIV) standard [?], [?] which can use for *6DoF* video compression. It uses multi-view RGB-D video as the data representation and includes the integrated pipeline for encoding, decoding,

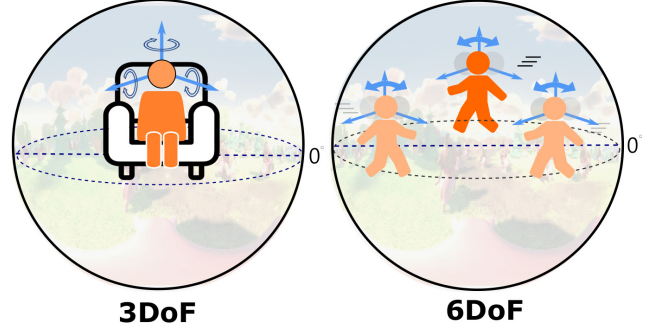


Fig. 1. Difference between *3DoF* and *6DoF* interactions.

synthesizing, and rendering. The Test Model for Immersive Video (TMIV) [?], [?], which is the reference software of MIV standard, has been released to show a reference implementation of MIV.

Besides, reducing bandwidth and maintaining high view quality is a bottleneck of *6DoF* real-time streaming. Because of the limitation of bandwidth, performing the best quality by adaptive quality model for *6DoF* immersive video is quite a challenging work. There are many existing rate control algorithms for video compression [?] [?] [?]. Based on the existed rate control algorithms, it is possible to optimize those models to predict the *6DoF* video performance. Although, there are still many effects that could impact the *6DoF* video quality, e.g., tile sizes [?], camera placements, complexity of scenes [?], number of groups, synthesizer, and a quantization parameter. In this paper, we conduct a rate-distortion model (RD-model) to predict the performance of *6DoF* immersive video in common situations as we show in Fig. 2. The model will design in an empirical way and aim at some vital TMIV parameters. The main goal of this model is to generate relevance between TMIV parameters and quality metrics, e.g., PSNR, SSIM, and VMAF. Once the model generates, it is possible to use on estimate the *6DoF* immersive video performance in different camera placements and also use it on real-time *6DoF* streaming.

II. RELATED WORK

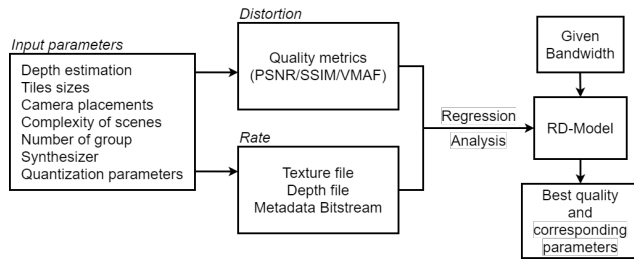


Fig. 2. RD-model workflow

A. 2D videos

B. 360 videos

C. Synthesized video

Yangang Cai [?] compressed the depth map by AVC, HEVC, and AVS3. Their results show that using the AVS3 encoder to compress the depth maps can provide better virtual view performance and less bitrate. Basel Salahieh [?] evaluate the performance of the object-based solution. Their results show the pixel rate saving and bitrate distortion in the object-based situation. Xavier Corbillon [?] implemented a 6DoF VR application with a multi-camera system and analyzed two extreme optimal algorithms. Their results show that tiling is able to improve the service performance and the high cost for the proactive optimizing strategies. By the previous experiments, both 6DoF video quality and bitrate have a negative correlation with QP.

REFERENCES

- [1] (2018) Augmented reality and virtual reality market. [Online]. Available: <https://reurl.cc/WLaNp9>
- [2] Meta. (2021) Meta. [Online]. Available: <https://about.facebook.com/meta/>