Tile-based VR Adaptive Streaming using JND Model

Your N. Here
Your Institution

Second Name
Second Institution

Abstract

1 Introduction

In recent years, the world is becoming more virtual than we ever thought it would be. Many video service providers, such as YouTube, roll out Virtual Reality (VR) videos which provide immersive experience to users. While consuming VR videos, users can change their viewpoint, resulting in an interactive experience than consuming traditional videos with a fixed viewing direction. However, VR videos' high demand of resolution and bitrates hinder their wide spread over the Internet.

The streaming of VR video is currently deployed in a naive way by streaming the entire video in constant quality, or simply building a cross-user viewpoint heat map and allocating higher quality to more important areas. However, when watching VR videos, users have their individual behaviors. Only a portion of the video is viewed by the user at a specific time, and different users interest in different areas[]. As a consequence, viewport-adaptive streaming is regarded as a promising way to deliver VR video through the Internet with limited bandwidth. One realization is tile-based streaming framework. In tile-based streaming, each temporal video segment is composed by several spatial tiles which can be independently encoded/decoded. Since each tile's Just-Noticeable Distortion (JND) is highly related to its distance from user's viewpoint, user chooses high bit-rate for near-viewpoint tiles, and chooses low bit-rate for far-viewpoint tiles.

But JND is not only related to viewpoint-content distance. Many prior works have proved that JND is at least related to luminance[], contrast[], complexity[]. We evaluate the potential improvement of these JND factors. When we take consideration of luminance, contrast and complexity of each tile and allocate bit-rate correctly, we can save x% bandwidth without decrease of user perceived quality. So there is big room for JND model to improve VR video streaming performance.

Although JND can save bandwidth significantly, applying JND model to tile-based streaming is challenging in two aspects:

- *User experience of VR video display is quite different from that of traditional non-VR video display in many aspects.* For example, in a VR video display, user has 360-degree view, immersive experience, content Depth of Field (DoF) awareness and active viewpoint. So JND model for VR display may be different from non-VR video display. Moreover, there might be some new JND factors which are never considered in traditional non-VR video display, which also .
- *Computing JND during VR video streaming is challenging.* Since server has information only about video contents, client-side bit-rate adaptation logic has information only about user viewpoint, neither of them can compute JND alone. How to make combination of these two informations and make bit-rate decision for each tile is a problem.

In this work, we address these challenges and present the design and implementation of a Tile-based VR Adaptive Streaming using JND Model (TAS-JND). TAS-JND is built on two key insights:

- *JND model in VR video display is very different from that in traditional non-VR video display.* On the one hand, current well-studied JND factors may influence JND differently in VR video and non-VR video. On the other hand, JND is also highly related to two important factors: viewpoint moving speed and content Depth of Field, which are never considered by previous JND model for non-VR videos. These two additional factors can further save us x% bandwidth without decreasing user perceived quality.
- **JND computing is difficult to be decoupled.** A simple idea of applying JND model to current adaptive streaming architecture is decoupling JND factors into video

content information and user behavior information. All JND factors about video content can be considered by video precoding. After that, user can compute JND and select bit-rate of each tile only based on user behavior (such as viewpoint position, viewpoint moving speed). However, the relationship between different JND factors are complex. According to our experiments, when we decouple JND factors, JND computing and rate selection will be very inaccurate, which greatly impairs user experience. So We need a new method to apply JND model to VR video adaptive streaming.

Taken together, these insights enable us to engineer a Tile-based VR Adaptive Streaming using JND Model. Firstly, in our JND model, luminance, contrast, complexity, viewpoint moving speed and content Depth of Field are taken into consideration. Secondly, a novel tile-based rate adaptation logic is presented to select bit-rate for each tile based on this JND model.

We implemented a prototype of TAS-JND and and integrated it in a VR video streaming system. (briefly describe our evaluation)

Contributions and Roadmap:

- Identifying key factors in building a JND model for VR video display. (§2-3)
- Presenting a tile-based VR adaptive streaming using JND model. (§4-5)
- Real-world evaluation that demonstrates substantial performance improvement by TAS-JND (§6).

2 Just-Noticeable-Distortion in VR Display

Since transmitting a VR-video needs huge bandwidth cost, researchers attend to cut video frames into number of tiles and allocate different quality to different tiles according to their importance, in order to achieve better perceived video quality with limited bandwidth.

Just Noticeable Distortion (JND) is a well-studied model to describe visual sensitiveness. It is widely used in video coding with the assumption that allocating low bit-rate to high-JND-area decreases less perceived video quality than allocating low bit-rate to low-JND-area.

As proved by many researches, JND is at least related to several factors, such as content lightness, contrast, distance to viewpoint,

However, VR video displays differ from non-VR video display in many aspects, such as 1, 2, 3, . So there are obviously some VR-only factors which are also related to JND. So far there is no prior works of building a VR-JND model for video streaming. So we aim to build a VR-JND model and design a VR video bit-rate adaptation based on it.

In section 2.1 we briefly introduce some JND factors which have been well-studied by prior works. In section 2.2 we present several VR-only JND factors along with our user studies.

2.1 Common-used JND factors

Brief introduce some JND factors and plot one or two most classical curve graphs.

2.2 VR-only JND factors

2.2.1 Vergence-accommodation conflict and visual threshold

Briefly introduce vergence distance and accommodation distance. (A figure is needed)

In non-VR video displays, the screen is bioptic, where only a single display is presented that is viewed by both eyes, so contents have no Depth of Field (DoF) information, vergence distance and accommodation distance are consistence. However, in most VR video displays, the screen is stereoscopic, where the illusion of depth is created by delivering images rendered from different angles to each eye. In this situation, accommodation distance is fixed but vergence distance is different for different contents, thus causes vergence-accommodation conflict.

Some prior works point out vergence-accommodation conflict can cause decrease of visual acuity. So we set up an experiment to measure the relationship between vergence-accommodation conflict and visual threshold.

Our experiment.

2.2.2 Light/Dark adaptation and visual threshold

Briefly introduce light/dark adaptation.

In non-VR video displays, the environment illumination totally depends on the real environment (e.g. under the sunlight, or in a classroom with electric lamp, or in a dark room). However, VR video displays are very different. When user wears HMD, the environment illumination totally depends on the content itself. So when the illumination changes dramatically, eyes need a period of time to adapt the new illumination.

Some prior works point out in the process of light/dark adaptation, visual acuity decreases. So we set up an experiment to measure the relationship between light/dark adaptation and visual threshold.

Our experiment.

2.2.3 Viewpoint moving speed and visual threshold

One of the most highlighted feature of VR video is that users can freely move their viewpoints. According to our data analysis, more than x% viewpoints are moving faster than y deg/s. (figure of our data analysis)

Some prior works point out when users viewpoint is moving, visual acuity decreases dramatically. So we set up an experiment to measure the relationship between viewpoint moving speed and visual threshold.

Our experiment.

3 JND Model

In this section we build VR-JND Model with consideration of 1, 2, 3, .

Structure of VR-JND Model. (a figure)

3.1 Formulation of factor 1

3.2 Formulation of factor 2

4 Tile-based VR Adaptive Streaming using JND Model

5 System Overview

6 Performance Evaluation

7 Related Work