Poster: A Hierarchical VR Streaming System through a WiFi Connection

Songzhou Yang[1], Junchen Guo[1], Xiaolong Zheng[2], Xinpeng Zhang[1], Chunya Liu[1], Pengyu Li[1], Meng Jin[1], Yuan He[1]
School of Software and BNRist, Tsinghua University[1]
School of Computer Science, Beijing University of Posts and Telecommunications[2]
{yangsz18, gjc16}@mails.tsinghua.edu.cn, zhengxiaolong.zxl@gmail.com, {zhangxp16, liuchuny17, lpy18}@mails.tsinghua.edu.cn, mengj@mail.tsinghua.edu.cn, heyuan@tsinghua.edu.cn

Abstract

Consumer Virtual Reality (VR) has been widely used in various application areas, such as entertainment and medicine. In spite of the superb immersion experience, to enable high-quality VR on untethered mobile devices remains an extremely challenging task. The high bandwidth demands of VR streaming generally overburden a conventional wireless connection, which affects the user experience and in turn limits the usability of VR in practice. In this work, we propose a hierarchical VR streaming through a WiFi connection. This design stems from the insight that humans field of view (FoV) is hierarchical, so that different areas in the FoV can be served with VR content of different qualities. By exploiting the gaze tracking capacity of the VR devices, this system is able to accurately predict the users attention so that the streaming of hierarchical VR can be appropriately scheduled. In this way, our work significantly reduces the bandwidth cost while keeping high quality of user experience.

1 Introduction

In spite of the rapid growth of VR market, there are huge gaps between the limited capacity of existing infrastructure and the high demand of VR streaming. First, a huge gap exists between the bandwidth capacity of conventional wireless technologies and the bandwidth demand of VR streaming. According to our calculation, the bandwidth demand is at least 4.2Gbps for 4K streaming with 120 frames per second (FPS). However, the bandwidth of the fastest commercial WiFi, 802.11ac, is 1.3 Gbps in theory and slower in practice. Second, the limited hardware resource on VR devices cannot meet the high computation demand of VR streaming. A tethered VR attached to a powerful PC may meet the demands, but the wires will restrict users even harm them.

This work proposes a hierarchical VR streaming via WiFi connection, which leverages the hierarchical property of human vision, predicts the user's attention, and accordingly schedules the mixed-quality VR streaming to the VR device.

2 Design

2.1 Overview

Figure 1 presents the framework of our design. Our system uses a C/S (client/server) architecture. The client consists of a VR HMD with video player and a attention prediction module. The server consists of a scheduling module and video resources from the providers.

2.2 Prediction of User Attention

In our system, we use the Fove VR as the VR HMD, which provides IMU based orientation tracking and IR-based position tracking. Besides, it also provides eye tracking with error less than 1°. We leverage the IMU sensors and the eye tracking results to further predict the user attention.

We adopt SVR (Support Vector Regression) [3] to perform the user attention prediction. We cannot pretrain a comprehensive model that satisfies all the conditions and achieves optimal performance in each specific condition at the same time. Therefore, in this work, both training and predicating are conducted online to adjust to the constantly changing user movement. We use a moving window that contains the movement data in recent 5 seconds to fit an attention transformation pattern. Then we leverage the pattern to predict the movement in the following 1 second.

2.3 Composition of Videos

Our system transmits mixed-quality video to the client for bandwidth consumption reduction, based on the user attention predication.

We first process the high bitrate videos offline into three types of definitions: HD (high-definition), SD (standard-definition), LD (low-definition).In addition, the latter two consist of multiple bitrates.

We then slice the video into many clips, each clip is 1 second long, consistent to the predication window length. If the length is too short, the real-time process will be too frequent which will result in wasting computing resource, and if the

International Conference on Embedded Wireless Systems and Networks (EWSN) 2019
25–27 February, Beijing, China

ISBN: 978-0-9949886-3-8

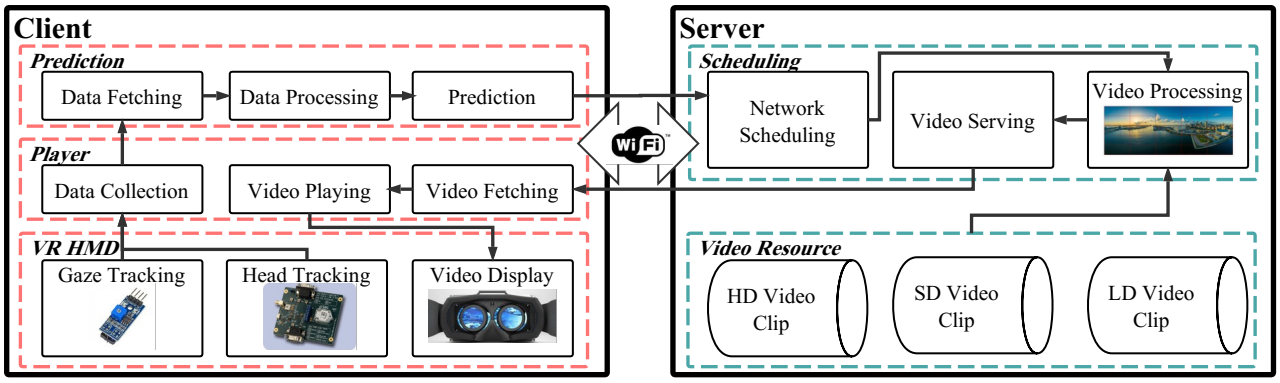

Figure 1. Overview of the hierarchical VR streaming

length is too long, we may mismatch the changeable network conditions so that users may gain a low QoE.

2.4 Streaming

The main task of the network scheduling module is to dynamically adjust some configurations in the system according to the current network condition, so that users can obtain the highest QoE while watching streaming.

We need a metric of QoE to formulate the optimization problem of our scheduling. MOS (Mean Opinion Score) [1] is the most common metric of QoE. But the MOS score is measured by a subjective quality evaluation test which we cannot easily get while streaming. But we can estimate it by the objective parameters, such as the bitrate. We use some datasets [4, 2] to explore the relationship between MOS and bitrate. We find that the relationship between bitrate and MOS score meet an exponential function, as shown in Equation 1, where the resolution is set as 114×60 (the resolution of one tile from 36×36 tiles for 4K resolution).

$$BVQA = \begin{cases} 1-e^{-0.648x} & \text{for LD areas;} \\ 1-e^{-0.324x} & \text{for SD areas;} \\ 1-e^{-0.081x} & \text{for HD areas.} \end{cases} \quad (1)$$

Now, we can define our QoE metric for a whole video clip as follow:

$$QoE = \sum_{i,j}^{i=N_x,j=N_y} BVQA_{ij} \times Weight_{ij} \quad (2)$$

where i, j is the index of tiles, N_x, N_y is the number of tiles in the vertical and horizontal direction, respectively.

Then, we formulate the scheduling problem as a Knapsack problem. Specifically, we have a set $\{x_1, x_2, ..., x_n\}$ of n video tiles, and each x_i is a set $\{d_1, d_2, ..., d_m\}$ of m kinds of definitions with weight w_{ij} referring to the bandwidth demand of video tile x_i with definition d_j. v_{ij} referring to the expected QoE of tile x_i with definition d_j which can be calculated by prediction results. We assume the whole available bandwidth as the weight capacity W. Then we formulate the optimization problem as follows.

$$\begin{aligned} \text{maximize} \quad & QoE = \textstyle\sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}x_i d_j \\ \text{subject to} \quad & \textstyle\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}x_i d_j \le W \\ & \text{and } x_i = 1 \text{ and } d_j \in \{0,1\} \\ & \text{and } \textstyle\sum_{k=1}^{m} d_k = 1 \end{aligned} \quad (3)$$

Note that, in the formulated problem, we always have $x_i = 1$ and $\sum_{k=1}^{m} d_k = 1$ to ensure integrity of video. We propose a greed algorithm to solve it. Finally, we compose videos according to streaming results.

3 Conclusions

In this work, we study the problem about how to enable high-quality VR streaming on untethered mobile devices without harming the QoE. The key challenge is the mismatch between the limited bandwidth capacity of existing wireless networks and the high bandwidth demands of high-quality VR streaming. In this work, we propose a practical VR streaming system that leverages the hierarchical human vision to lower the video qualities of the unneeded fields. By providing the mixed-quality video, our design significantly reduces the video size and the bandwidth demand as well. It also integrates an online attention prediction algorithm that leverages the head and gaze movements to predict the user attention field where the high quality video should be displayed. Based on the predicted user attention, our design further schedules the the composition and transmission of videos to reduce the service delay and improve the QoE.

4 Acknowledgments

This work was supported by National Natural Science Foundation of China under grant No. 61672240.

5 References

[1] Vocabulary for performance and quality of service. https://www.itu.int-/rec/T-REC-P.10.

[2] M. Cheon and J.-S. Lee. Subjective and objective quality assessment of compressed 4k uhd videos for immersive experience. *IEEE TCSVT*, 28, 2018.

[3] A. J. Smola and B. Scholkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3), 2004.

[4] Y. Zhu, L. Song, R. Xie, and W. Zhang. Sjtu 4k video subjective quality dataset for content adaptive bit rate estimation without encoding. In *Proceedings of IEEE BMSB*, 2016.