Research Proposal: 6-DoF Immersive Video Streaming to Head-Mounted Displays

Tse-Hou Hung
tsehou.nthu@gmail.com
National Tsing Hua University
Hsinchu, Taiwan

ABSTRACT

Six Degree-of-Freedom (6DoF) Immersive video streaming is very challenging due to various data representations, strict network condition and computing requirements ,and complex user experience models. In my PhD study, we study on three research problems to overcome these three challenges. First, we study different data representations to understand their pros, cons, and suitable usage scenarios. Second, we optimize the streaming systems to reduce the bandwidth requirements in immersive video streaming. Last, we design and conduct a series of user studies to derive the user experience. The results of these three steps lead to a fully-optimized immersive video streaming systems, which are crucial to many novel applications, including holographic conferences, real scene streaming, remote surgery, and fire fighter (or military) training.

1 INTRODUCTION

Virtual Reality (VR) technology is thriving in various business sectors, including computer games, tourism industry, real estates, and occupational trainings. The VR scenes can either be generated with computer graphics or captured from nature scenes, which provide omnidirectional, a.k.a. 360°, viewing experience of virtual worlds. It is estimated that the global VR market size will reach 26.89 billion USD by 2022, with an annual growth rate of 54% from 2017 to 2022 [10]. As VR technology becomes more and more mature, researchers start to study its Quality of Experience (QoE) in order to deliver higher-quality scenes for better user experience.

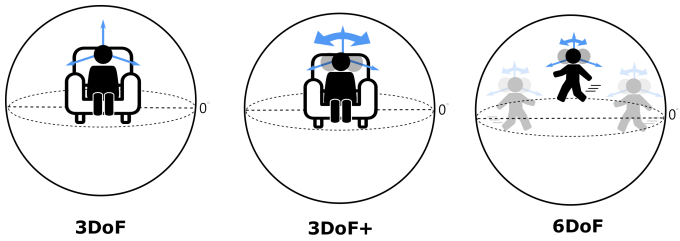

Figure 1: Three phases of VR technology development. The spheres represent the virtual worlds.

One key factor affecting the QoE of VR technology is the way users interact with the virtual worlds. MPEG-I group has defined three development phases of the VR technology: 3 Degree-of-Freedom (3DoF), 3DoF+, and 6DoF, which are illustrated in Fig. 1. 3DoF VR technology allows users to rotate their heads in three dimensions, which are represented by yaw ψ, pitch θ, and roll ϕ. However, with 3DoF VR, when users move their heads or stand up and walk, their position changes do *not* affect the views rendered in the Head-Mounted Displays (HMDs). To overcome this limitation, 3DoF+ VR is proposed to support *limited* movements. Namely, users can slightly move their heads to a certain extent when sitting on fixed chairs. 6DoF VR further enables users to freely move, e.g., walk, in virtual worlds. Users can not only rotate their heads in three dimensions but also move in the three additional dimensions, which are x, y, z coordinates.

A naive way to achieve 3DoF+ and 6DoF is to capture multiple 360° videos at different positions [2, 9], and only allow the HMD users to switch among these discrete positions in the scenes. We target more general 3DoF+ and 6DoF VR, where users can freely move in among the discrete positions. We refer to the streaming systems that support such 3DoF+ and 6DoF VR as *immersive video streaming* throughout this proposal.

Achieving real immersive video streaming is not an easy task, because there are many challenges that need to be overcome:

- Different Data Representations: Various data types can be used to store the information of immersive video, e.g., 2D texture and depth image, light field, 3D point cloud, and 3D mesh. Each of them has pros, cons, and usage scenarios, which have not been comprehensively studied.
- Strict Network Condition Requirements: The data representations of immersive video have a huge data size. They need a lot of bandwidth to be streamed, while the realtimeness needs to be maintained. Take light field video as an example, streaming it consumes bandwidth in the range between 200 Gb/s and 1 Tb/s [1], which is much higher than the available bandwidth we have today. Moreover, the interactive applications, e.g., 6DoF VR games and holographic conferences, need low latency to realize real-time communications. This in turn makes the network condition requirements more strict.
- Complex User Experience Models: User experience is the quality perceived by users. It is an important performance metric for multimedia applications. However, the user experience of immersive videos is still not explored, and the challenges of measuring user experience include but not limited to: (i) too many parameters may affect the user experience in immersive video streaming and (ii) it's hard to build an immersive video streaming testbed.

In this research proposal, we propose three research directions to address the three challenges: (i) *Data Representations* (ii) *Optimal Streaming* (iii) *Quality of Experience*. We introduce them in details below.

2 RESEARCH PROBLEMS

2.1 Data Representations

Whiles data types mentioned in Sec. 1 could be suitable to immersive video streaming, Their pros and cons are unknown. In our research group, we have studied 3DoF VR streaming in the past few years [4, 5]. These works help us to understand the existing streaming techniques in 3DoF 360° video streaming. We are currently expanding our research to different 6DoF data types to understand their pros and cons. We will also concretize their suitable usage scenarios. The final outcome of this work is several immersive video streaming testbeds that support heterogeneous data representations. Through real experiments, we get to know more systems challenges in these emerging applications.

2.2 Optimal Streaming

To reduce the bandwidth and computing requirements of immersive video streaming, the streaming systems must be optimized. In our recent work [7], we adopt Test Model for Immersive Video (TMIV) [11–13], which is a Depth Image Based Rendering (DIBR) codec from MPEG, to optimize the immersive video streaming. We develop algorithms to solve the configuration optimization problem based on deep learning, particularly the Neural Network (NN) approaches. The two proposed algorithms are: a Convolutional Neural Network (CNN) based algorithm and a Deep Reinforcement Learning (DRL) based algorithm Our CNN algorithm benefits from: (i) automatic extracting latent features and (ii) inferencing the prediction rapidly and directly according to the input, which allows the configuration optimizer to adapt to various video scenes and dynamic camera parameters. Our DRL algorithm systematically builds an *agent* that can adapt to dynamic *environments* [3, 6, 9, 14]. The trained agent learns how to quickly search through a large space for optimal configurations. The two proposed algorithms are trained to adaptively find the optimal configurations given the diverse video content, HMD user behaviors, and user-specified utility functions. We submitted our paper to ACM Multimedia Conference 2020, which is under review.

We are currently extending this work into a journal submission. In particularly, we plan to (i) design, implement, and evaluate an end-to-end DIBR-based immersive video streaming system, (ii) apply the state-of-the-art DRL algorithms and (iii) adopt larger datasets to train our optimizers. We expect the performance of our systems can be improved after we apply these optimization techniques. Moreover, according to the research outcomes of Sec. 2.1, we will generalize our solutions to other data types/representations.

2.3 Quality of Experience (QoE)

To measure QoE of users, we plan to design and conduct a series of user studies to quantify the relationship between each parameters and user experience in immersive video streaming sessions. In our previous work [7], we conduct a small-scale user study to evaluate our solution. The testbed used in that work can be extended and used in more comprehensive subjective experiments. The high-level architecture of the immersive video streaming system is shown in Fig. 2. We also plan to measure the Just-Noticeable Difference (JND) [8] bitrate of immersive video streaming. Using JND bitrates, we can intelligently save the network and computing resources without affecting the user experience. Combining the outcomes of the QoE study with the ones achieved in Sec. 2.1 and Sec. 2.2 gives a fully-optimized immersive 6DoF video streaming system.

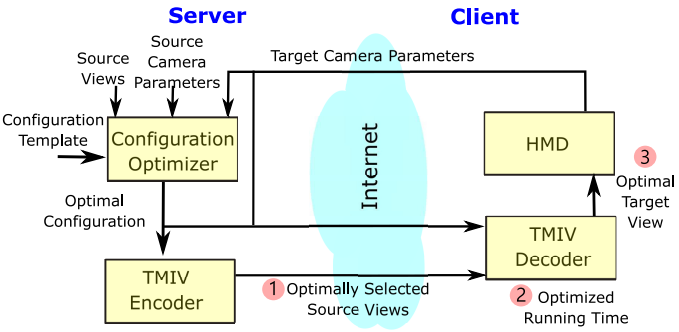

Figure 2: High-level architecture of immersive video streaming systems for our QoE study.

3 RESEARCH PLAN

Fig. 3 gives the Gantt chart of my research plan. The first problem is the data representation study. The outcome of this research problem is our better understanding on the immersive video streaming systems. That is also useful to the research community. Concurrently, we will extend our recent work [7] into a journal paper. After that, we work on the QoE study. We plan to build an immersive video streaming testbed and conduct a series of user studies to understand the user experience. We will also continue optimizing the immersive video streaming testbed by applying different optimization algorithms developed by us. The results of QoE study can also help us better optimize the testbed. Last, we will design, build,

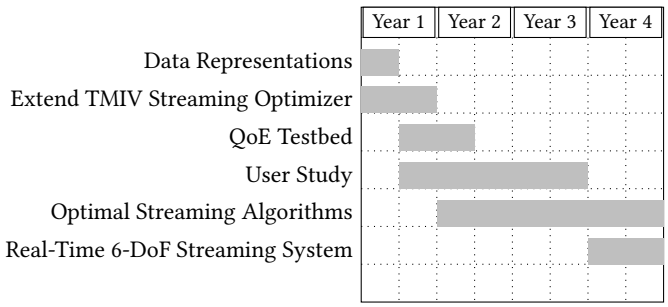

Figure 3: Gantt chart of my PhD study.

and evaluate a real-time immersive video streaming system based on the results of our research. The resulting optimized testbed will be useful to many novel applications.

4 EXPECTED IMPACTS

Currently, the techniques of immersive video streaming are still in their infancy. Most of the challenges identified in this proposal have not been rigorously studied. Indeed, existing VR/AR applications do not support real-time streaming. Even if they do, only 3DoF interactions are supported, which is not the *real* immersive experience.

In my PhD study, we will propose a series of solutions to overcome the challenges in immersive video streaming. As system researchers, we will evaluate these solutions through actual experiments. In data representation work, we will provide a comprehensive review, which can help the research community gain more knowledge on the features, challenges, and opportunities of various 6DoF data types. Through user study and data analysis, we will find new discoveries on the user experience of immersive video streaming in the QoE work. With these discoveries, the content creators can better meet users' needs when they create the immersive content. The researchers can also use the users' viewing behaviors to develop more efficient streaming solutions. In optimal streaming, we will propose innovative algorithms to overcome various issues caused by scarce resources. The bandwidth and latency requirements of immersive video streaming will, therefore, be fulfilled under diverse network conditions, and several 6DoF applications can thus be realized.

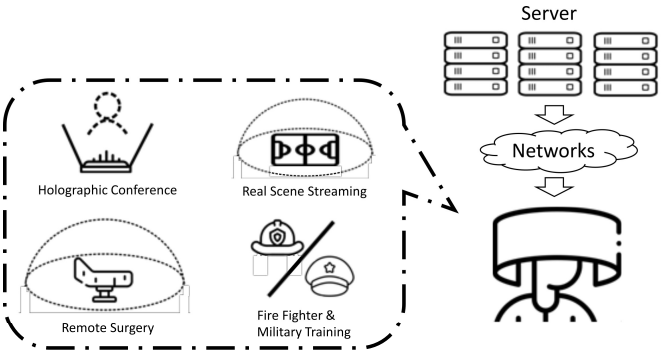

Figure 4: Usage scenario of immersive video streaming.

The techniques of immersive video streaming can make people's life more convenient and productive. As shown in Fig. 4, immersive video streaming can be used in various usage scenarios, such as:

- **Holographic conferences** may become reality. Users can see the projection of remote people and naturally communicate with one another . The holographic conferences will provide exactly the same experience as face-by-face ones.
- **Real scene streaming** may provide immersive live streaming for sports, speech, and other events. For example, sports fans can watch live sports without blind angles to enjoy the games.
- **Remote surgery** can also be performed. Immersive video streaming can provide the situations of remote patients to the doctor, who can carry out the surgery remotely. This will improve the healthcare quality in rural areas.
- **Fire fighter and military training** can be done in a safer and more efficient way. The training sessions can simulate real situations to help learners experience danger environments without risks.

We note that these usage scenarios are just the representative ones. Many other applications of immersive video streaming are yet to discovered in upcoming years.

REFERENCES

[1] A. Clemm, M. Vega, H. Ravuri, T. Wauters, and F. Turck. 2020. Toward Truly Immersive Holographic-Type Communication: Challenges and Solutions. *IEEE Communications Magazine* 58, 1 (2020), 93–99.

[2] X. Corbillon, F. Simone, G. Simon, and P. Frossard. 2018. Dynamic Adaptive Streaming for Multi-viewpoint Omnidirectional Videos. In *Proc. of ACM Multimedia Systems Conference (MMSys'18)*. 237–249.

[3] L. Costero, A. Iranfar, M. Zapater, F. Igual, K. Olcoz, and D. Atienza. 2019. MAMUT: Multi-Agent Reinforcement Learning for Efficient Real-Time Multi-User Video Transcoding. In *Proc. of IEEE Design, Automation Test in Europe Conference Exhibition (DATE'19)*. 558–563.

[4] C. Fan, W. Lo, Y. Pai, and C. Hsu. 2019. A Survey on 360° Video Streaming: Acquisition, Transmission, and Display. *Comput. Surveys* 52, 4 (2019).

[5] C. Fan, S. Yen, C. Huang, and C. Hsu. 2020. Optimizing Fixation Prediction Using Recurrent Neural Networks for 360° Video Streaming in Head-Mounted Virtual Reality. *IEEE Transactions on Multimedia* 22, 3 (2020), 744–759.

[6] T. Huang, R. Zhang, C. Zhou, and L. Sun. 2018. QARC: Video Quality Aware Rate Control for Real-Time Video Streaming Based on Deep Reinforcement Learning. In *Proc. of ACM International Conference on Multimedia (MM'18)*. 1208–1216.

[7] T. Hung, C. Hsu, and C. Hsu. 2020. *Technical Report: Optimizing Immersive Video Streaming Using Deep Learning Approaches: A Case Study on TMIV*. Technical Report. https://reurl.cc/z8YVbe.

[8] Y. Jia, W. Lin, and A. A. Kassim. 2006. Estimating Just-Noticeable Distortion for Video. *IEEE Transactions on Circuits and Systems for Video Technology* 16, 7 (2006), 820–829.

[9] H. Pang, C. Zhang, F. Wang, J. Liu, and L. Sun. 2019. Towards Low Latency Multi-viewpoint 360° Interactive Video: A Multimodal Deep Reinforcement Learning Approach. In *Proc. of IEEE Conference on Computer Communications (INFOCOM'19)*. 991–999.

[10] ZION Market Research. 2018. Virtual Reality (VR) Market by Hardware and Software for (Consumer, Commercial, Enterprise, Medical, Aerospace and Defense, Automotive, Energy and Others): Global Industry Perspective, Comprehensive Analysis and Forecast, 2016–2022. Retrieved April 21, 2020 from https://www.zionmarketresearch.com/report/virtual-reality-market

[11] B. Salahieh, B. Kroon, J. Jung, and M. Domański. 2019. Test Model 2 for Immersive Video. International Organization for Standardization Meeting Document ISO/IEC JTC1/SC29/WG11 MPEG/N18577.

[12] B. Salahieh, B. Kroon, J. Jung, and M. Domański. 2019. Test Model 3 for Immersive Video. International Organization for Standardization Meeting Document ISO/IEC JTC1/SC29/WG11 MPEG/N18795.

[13] B. Salahieh, B. Kroon, J. Jung, and M. Domański. 2019. Test Model for Immersive Video. International Organization for Standardization Meeting Document ISO/IEC JTC1/SC29/WG11 MPEG/N18470.

[14] R. Sutton and A. Barto. 2018. *Reinforcement Learning: An Introduction* (2 ed.). A Bradford Book.