Checkpoint 1

作者為 元駿 孫

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] or Google [3], [4].

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 [5], [6] which can use for 6DoF video compression. It uses multi-view RGB-D video as the data representation

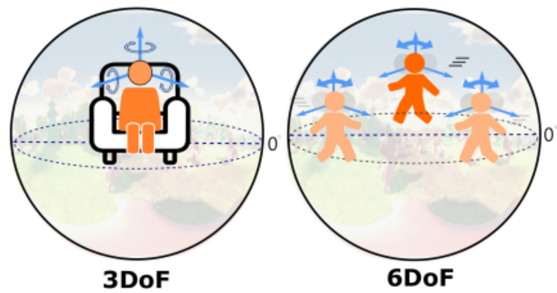

Fig. 1. Difference between 3DoF and 6DoF interactions.

and includes the integrated pipeline for encoding, decoding, synthesizing, and rendering. The Test Model for Immersive Video (TMIV) [7], [8], 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 [9] [10] [11]. 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 [12], camera placements, complexity of scenes [13], 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., PSRN, 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.

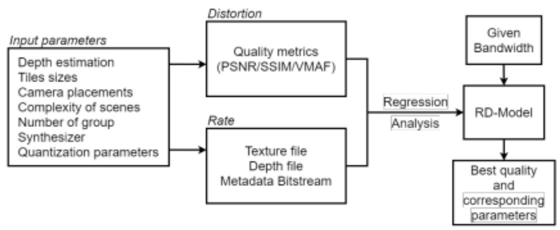

Fig. 2. RD-model workflow

II. Related Work

A. 2D videos

B. 360 videos

C. Synthesized video

Yangang Cai [14] 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 [13] 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 [15] 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.

III. MPEG Immersive Video Standard

In this section, we briefly introduce the workflow and components of MIV codec [5], [7].

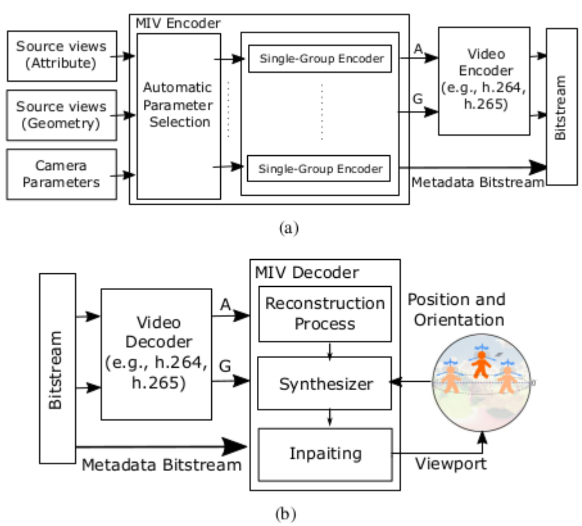

Fig. 3. The high-level overview of process flow of TMIV: (a) Encoder and (b) Decoder.

Fig. 4. The example of atlases. The left picture is attribute atlas, and the right picture is geometric altas.

Fig. 3(a) shows the high-level workflow of MIV encoder. The inputs of MIV encoder are *source views*. Each source view is composed of attribute (texture) videos, geometric (depth) videos, and camera parameters. MIV encoder do the following process to compress source views:

- **Automatic parameter selection.** MIV encoder automatically calculate the parameters for compression, e.g., assessing geometric video quality, splitting source views into multiple group according to configuration, and labeling source views in each group.
- **Single-group encoders.** MIV encoder encodes each group of source views separately. In each group, the encoder chooses several views as the basic view according to the label of source view, and remove the duplicate area in other source views. The basic view and remaining area of other views are packed into rectangle video frames, which are called *atlases*. Fig. 4 show the example of atlases.

The outputs of MIV encoder are attribute atlases, geometric atlases, and metadata bitstream. The atlases are further compressed by video codec, and multiplexed with metadata bitstream as a single bitstream.

Fig. 3(b) shows the high-level workflow of MIV decoder. The inputs pf MIV decoder is the bitstream contains atlases bitstream, and metadata bitstream. The video decoder first be employed to decompress attribute atlases and geometric atlases. After that MIV decoder do the following process to decompress altases and synthesize the user's viewport.

- **Reconstruction process.** The MIV decoder reconstruct the source view by using the data in atlases.
- **Synthesizer.** The MIV decoder employ view synthesis techniques to synthesize the user's viewport according to user's position and orientation. Specifically, the synthesizer warp the pixel of each source view to user's viewport according to depth information, and blending the pixel values from each source views.
- **Inpainting.** After synthesis, the synthesized result may contain holes without information. The inpainting process

uses the information from neighbor pixels to calculate pixel value for holes.

The outputs of MIV decoder are user's viewport synthesized according to user's position and orientation.

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/

[3] Google. (2021) Google ar and vr. [Online]. Available: https://arvr.google.com/

[4] M. Broxton, J. Flynn, R. Overbeck, D. Erickson, P. Hedman, M. DuVall, J. Dourgarian, J. Busch, M. Whalen, and P. Debevec, "Immersive light field video with a layered mesh representation," vol. 39, no. 4, pp. 86:1–86:15, 2020.

[5] J. M. Boyce, R. Doré, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V. K. M. Vadakital, and L. Yu, "Mpeg immersive video coding standard," *Proceedings of the IEEE*, 2021.

[6] MPEG. (2021) Mpeg immersive video. [Online]. Available: https://mpeg.chiariglione.org/standards/mpeg-i/immersive-video

[7] B. Salahieh, J. Jung, and A. Dziembowski, "Test Model 10 for MPEG Immersive Video," International Organization for Standardization Meeting Document ISO/IEC JTC1/SC29/WG04 N0112, 2021.

[8] MPEG. (2019) The gitlab of mpeg test model for immersive video. [Online]. Available: https://gitlab.com/mpeg-i-visual/tmiv/-/tree/v10.0.1

[9] M. A. Papadopoulos, F. Zhang, D. Agrafiotis, and D. Bull, "An adaptive qp offset determination method for hevc," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 4220–4224.

[10] Z. Wu, H. Yu, B. Tang, and C. W. Chen, "Adaptive initial quantization parameter determination for h.264/avc video transcoding," *IEEE Transactions on Broadcasting*, vol. 58, no. 2, pp. 277–284, 2012.

[11] Z. He and S. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, pp. 1221–1236, 2001.

[12] J.-B. Jeong, S. Lee, I.-W. Ryu, T. T. Le, and E.-S. Ryu, "Towards viewport-dependent 6dof 360 video tiled streaming for virtual reality systems," in *Proceedings of the 28th ACM International Conference on Multimedia*, ser. MM '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 3687–3695. [Online]. Available: https://doi.org/10.1145/3394171.3413712

[13] B. Salahieh, W. Cochran, and J. Boyce, "Delivering object-based immersive video experiences," pp. 103–1–103–8(8), 2021.

[14] Y. Cai, R. Wang, K. Qiu, R. Peng, Z. Cheng, and Q. Wang, "Depth map video compression performance evaluation for ieee 1857.9," in *2021 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2021, pp. 1–6.

[15] X. Corbillon, F. De Simone, G. Simon, and P. Frossard, "Dynamic adaptive streaming for multi-viewpoint omnidirectional videos," in *Proceedings of the 9th ACM Multimedia Systems Conference*, ser. MMSys '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 237–249. [Online]. Available: https://doi.org/10.1145/3204949.3204968

Checkpoint 1

原創性報告

相似度指數	網際網絡來源	出版物	學生文稿
23%	20%	18%	10%

主要來源

1	Submitted to University of North Florida 學生文稿	4%
2	trepo.tuni.fi 網際網絡來源	2%
3	www.koreascience.or.kr 網際網絡來源	2%
4	etd.lib.nsysu.edu.tw 網際網絡來源	2%
5	www4.comp.polyu.edu.hk 網際網絡來源	2%
6	Submitted to Cornell University 學生文稿	2%
7	Daniel J Ringis, Francois Pitie, Anil Kokaram. "Near Optimal Per-Clip Lagrangian Multiplier Prediction in HEVC", 2021 Picture Coding Symposium (PCS), 2021 出版物	2%
8	Patrick Garus, Felix Henry, Joel Jung, Thomas Maugey, Christine Guillemot. "Immersive	2%

Video Coding: Should Geometry Information be Transmitted as Depth Maps?", IEEE Transactions on Circuits and Systems for Video Technology, 2021
出版物

9	Submitted to University College Falmouth 學生文稿	1 %
10	www.dis.cwi.nl 網際網絡來源	1 %
11	www.sciencegate.app 網際網絡來源	1 %
12	Tuan Thanh Le, Jong-Beom Jeong, SangSoon Lee, Jaehyoun Kim, Eun-Seok Ryu. "An Efficient Viewport-Dependent 360 VR System Based on Adaptive Tiled Streaming", Computers, Materials & Continua, 2021 出版物	1 %
13	Basel Salahieh, Mengyu Chen, Jill Boyce. "An Overview of MPEG Immersive Video", OSA Imaging and Applied Optics Congress 2021 (3D, COSI, DH, ISA, pcAOP), 2021 出版物	1 %
14	Bin Wang, Yule Sun, Lu Yu. "Low Pixel Rate 3DoF+ Video Compression Via Unpredictable Region Cropping", 2019 Picture Coding Symposium (PCS), 2019 出版物	1 %

15 Yangang Cai, Ronggang Wang, Ke Qiu, Rui Peng, Zhipeng Cheng, Qi Wang. "Depth Map Video Compression Performance Evaluation For Ieee 1857.9", 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2021 <1 %

出版物

排除引述	關閉	排除相符處	關閉
排除參考書目	關閉		