Signal Processing 147 (2018) 154–162

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

JET: Joint source and channel coding for error resilient virtual reality video wireless transmission

Zhi Liu[a,*], Susumu Ishihara[a], Ying Cui[b], Yusheng Ji[c], Yoshiaki Tanaka[d]

[a] *Shizuoka University, 3-5-1 Johoku, Hamamatsu 432–8561, Japan*
[b] *Shanghai Jiao Tong University, 800 Dongchuan Rd, Minhang Qu 200240, China*
[c] *National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101–8430, Japan*
[d] *Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169–8555, Japan*

ARTICLE INFO

Article history:
Received 1 August 2017
Revised 16 December 2017
Accepted 4 January 2018

Keywords:
Virtual reality
VR
360 video
Source and channel coding
Video streaming
Error resilient

ABSTRACT

Virtual reality (VR) provides users extraordinary viewing experience and draws more and more attentions from both industry and academia. In this paper, we propose JET: *Joint source and channel coding for Error resilient virtual reality video wireless Transmission*, where we jointly investigate how to conquer the problem of source video's huge size, how to efficiently satisfy a user's view switch request and how to handle packet loss. Specifically, we first divide a VR video into smaller video tiles. Upon a user's view switch request, the tiles corresponding to the part that the user is requesting, referred to as the *field of view* (FoV), and the part that the user may switch to before new video can be received, are delivered over a wireless network. We consider *unequal error protection* (UEP) for FoV and the rest part and formulate the inherent error resilient VR video transmission problem into a joint source and channel coding problem. In particular, we optimize the tile partition, quantization parameter and *Forward Error Correction* (FEC) packet allocation to maximize a user's received video quality. We also propose a low-complexity heuristic algorithm to solve this optimization problem. Extensive simulations are conducted and simulation results verify the superior performance.

1. Introduction

Virtual reality (VR) [1] video can provide users extraordinary viewing experience and has drawn great attentions in both industry and academia. A great amount of effort has been devoted to the researches and developments of VR devices (such as *HTC Vive* and *Sony Playstation VR*), VR content generation (i.e. how to capture a scene using multiple cameras) and VR video encoding (i.e. how to encode the VR video efficiently [2]), etc. Due to its features, VR has been widely used in many fields [3,4] such as training, education and entertainment, and augmented reality (AR)/VR revenue is expected to reach 150 billion USD by 2020 [5].

To realize and enhance VR related applications, transmitting a VR video to a user over a lossy wireless channel is essential. VR video streaming has thus become a research hotspot [6,7]. One major research issue of video streaming is how to handle the unavoidable packet loss over wireless networks, which degrades the transmission performance. Traditional error resilient video streaming has been widely studied in many aspects, such as joint source and channel coding [8,9], error resilient video encoding [10], multi-path transmission [11], lost frame recovery [12], scalable video streaming [13–15], cooperative video streaming and recovery [16–19], video streaming over various types of networks [20–23] and other issues [24–26]. However, the existing results for traditional video streaming cannot be directly applied to VR video streaming. This is because the traditional streaming method does not consider VR's unique features such as VR video's huge size and view switch requirements.

In most of the VR video streaming researches, a common method for handling the large size of a VR video is *divide*, i.e. given that a user only watches a part of a VR video each time, a VR video can be partitioned into smaller pieces, and only the part corresponding to the user's gaze point (i.e. what user is watching) is sent to the user. The video piece (or segment) is named tiles in this paper, and we use these terms interchangeably throughout the paper. In this way, the transmission load gets greatly reduced and the user's requirement may still be satisfied. Specifically, El-Ganainy and Hefeeda [27] investigates several aspects in building a VR video streaming system. Hosseini and Swaminathan [28], Hosseini [29] propose a bandwidth efficient field-of-view-aware streaming method by exploiting the semantic link of video

* Corresponding author.
E-mail addresses: liu@shizuoka.ac.jp, liu@ieee.org (Z. Liu).

https://doi.org/10.1016/j.sigpro.2018.01.009

pieces and a user's viewpoint. Ghosh et al. [30] proposes to fetch the invisible portion (i.e. the part that user is not watching) at the lowest quality and adaptively decides the quality for the part what the user is watching based on the bandwidth prediction. Qian et al. [31] conducts a measurement study on the commercial 360 video platforms, and proposes a cellular-friendly streaming scheme that delivers only 360 videos' visible portion based on head movement prediction. In [32], Oculus 360 degree video streaming is studied and some insights from measurements are discussed. A resource management method for VR in heterogeneous small cell networks is studied in [33]. TaghaviNasrabadi et al. [34] proposes a layer solution for the 360 video over the networks to improve the *quality of experience* (QoE) by reducing the probability of video freezes and the response latency. The authors in [35] propose to transmit the part being watched and the part possibly to be switched within a *round-trip time* (RTT), to solve the problem raised by the huge size of a VR video. The quantization parameters are optimized to maximize users' received video quality.

Different from these related schemes, we take the large-size VR video's large size and view switch requirements into consideration and focus on the error resilient VR video streaming over lossy wireless networks. In this paper, we propose *JET: Joint source and channel coding for Error resilient virtual reality video wireless Transmission*, a *field-of-view* (FoV) (i.e. FoV) based VR streaming solution, which jointly investigates how to conquer the problem of the huge VR source video, how to provide users with view switch abilities and how to provide error resilient VR video streaming. Specifically, we first divide the VR video into smaller video tiles, each tile can be encoded using the state-of-the-art encoding scheme such as HEVC or H.264. Upon a user's view switch request, a video stream will be delivered to the user via a wireless network. The delivered video stream includes the tiles of the FoV and the part of the video that the user may switch to before new video can be received (i.e. during the time period equals to RTT). In other words, zero view switch delay is provided when the user switches to other viewing angles. We use Gilbert-Elliot model to model the channel loss [36,37] and consider FEC to protect channel loss. We consider *unequal error protection* (UEP) for FoV and the rest part, and formulate the inherent error resilient VR video transmission problem into a joint source and channel coding problem. In particular, we optimize the tile partition, quantization parameter for each tile and *Forward Error Correction* (FEC) packet allocation to maximize user's received video quality. We then discuss the optimal solution and propose to obtain a promising solution. Extensive simulations are conducted and the simulation results show that the proposed scheme can outperform comparison schemes in typical network scenarios.

The rest of this paper is organized as follows: Section 2 introduces the VR video streaming system. In Section 3, we formulate the VR video streaming problem into a source and channel coding problem, and propose a heuristic algorithm to optimize the problem. The simulation setups and results are presented in Section 4. Section 5 concludes the paper.

2. VR video streaming system

In this section, we overview the VR video streaming system and introduce the FoV-based streaming solution, view switch model and channel loss model.

2.1. System overview

In the proposed system *JET*, there is one base station, one user and one VR video. Fig. 1 illustrates such a VR video streaming system, which mainly involves the content generation, encoding, transmission and decoding. As shown in Fig. 1, the scene of interest is captured by multiple cameras and the server stores the generated VR video for the user to subscribe. The time is quantized to be discrete. At each time instance, the user has the freedom to select the viewing angle to watch and has the freedom to switch to another viewing angle during the VR video playback. Ideally, an infinity number of viewing angles should be provided, which makes the streaming difficult. In this paper, we limit the number of possible viewing angles to be finite, the details of which will be discussed later. By using relatively a large number of the possible viewing angles, the viewing experience is close to the case of infinite viewing angles, and the implementation complexity is greatly reduced.

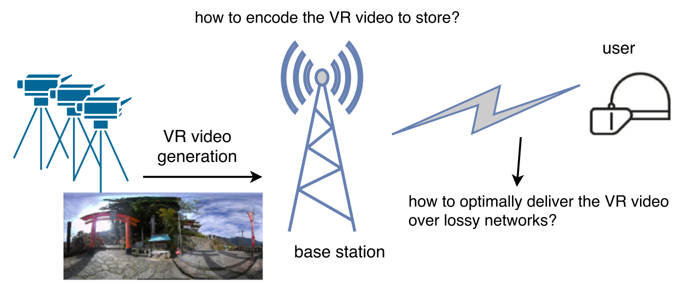

Fig. 1. Illustration of the VR transmission system, which mainly involves the content generation, encoding, transmission and decoding.

Sending the whole VR video to the user is extremely challenging over the current wireless network (such as Wi-Fi and cellular networks) due to the large size of a VR video, limited bandwidth, lossy nature of wireless channels and viewing angle switch requirements. Since each time the user only watches one part of the VR video sphere, the VR video is divided into smaller segments for the user to subscribe. Upon the user's VR video subscription request, the server sends the corresponding VR video stream, which covers the user's FoV and the part of the video that user may switch to before a new requested stream is delivered to the user. The FoV is what the user is watching currently as indicated by the user's view request. Note that each time the delivered video stream is composed of a number of tiles, and this number is determined by the size of the FoV, location of the FoV and size of each tile. Once receiving and decoding the video stream, the user can playback the VR video and switch viewing angles during the video playback with zero switch delay. Note that the view switch with zero delay is guaranteed by the VR video stream construction from tiles. The details will be discussed in the following. Retransmission is not adopted in this VR video streaming system, as it may result in a large delay. FEC packets are used to protect the transmission over the lossy channel.

2.2. FoV based streaming solution

The interaction between the server and the user is not instant. There is a time interval between the time epoch when the user sends a new view request and time epoch when the user receives the requested video from the server. The uploading of the user's viewing request to the server and downloading the user's requested video needs certain time, which is named as *RTT* in the literature. *RTT* may be a variable, and is usually at hundred milliseconds level at most. We consider the largest *RTT* T_0 and guarantee that the view switch request during *RTT* can be satisfied. In particular, given that the user can switch viewing angle but cannot change the stream during T_0, an extra part of the video besides the video part corresponding to the FoV has to be delivered, to guarantee that the user's view switch request can be satisfied with zero delay. The number of view switches during T_0 is decided by