Signal Processing 145 (2018) 285–294

Contents lists available at ScienceDirect

Signal Processing

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

A packetization strategy for interactive multiview video streaming over lossy networks

Xue Zhang[a], Yao Zhao[a,*], Tammam Tillo[b], Chunyu Lin[a]

[a] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
[b] The Faculty of Computer Science, Free University of Bozen-Bolzano, Bolzano BZ 39100, Italy

ARTICLE INFO

Article history:
Received 30 July 2017
Revised 1 December 2017
Accepted 21 December 2017
Available online 23 December 2017

Keywords:
Interactive multiview video streaming
Multi-view video plus depth
Depth-image-based rendering
Packetization
Error resilience

ABSTRACT

Interactive multiview video streaming (IMVS) applications permit to freely navigate within 3D scenes. This endows end-users the immersive and virtual reality experiences. Any additional viewpoint of interest can be virtually synthesized by received texture views with associated depth maps. In this paper, we consider the setup where both the encoded texture and depth bitstreams experience packet losses during network streaming. We investigate two packetization strategies and develop a novel strategy to improve error-resilience capabilities for multiview video plus depth transmission, where texture data and the corresponding depth go into the same packet. We examine the resulting distortion of different packetization schemes and analyze the real difference between them in diverse scenarios. In particular, we take into account: (i) various packet loss patterns, where the effect of a burst loss is identified as an important characteristic; (ii) two packet loss models that characterize different probabilities for each scenario. Experimental results demonstrate that, compared with the competing approach, our proposed packetization scheme increases the visual quality by up to 2.32 dB and 1.55 dB for texture views and synthesized views, respectively.

1. Introduction

Recent advances in 3D video capturing and communication technologies as well as virtual reality, pushed ahead the research for novel user-centric multimedia paradigms. Interactive multiview video streaming (IMVS) [1] is one typical application which provides end-users the possibility to navigate within a 3D scene in real-time fashion. Thus, it provides an immersive experience of real-world. However, to enable such interactivity, the multiview video content needs to be captured by an array of spaced cameras to cover different viewing angles. To remove the intrinsic redundancy from such data, interview prediction, in addition to traditional temporal prediction, is used in current state-of-the-art standards. For example, the multiview video coding (MVC) [2], multiview and 3D video extensions of high efficiency video coding (HEVC) standard namely MV-HEVC and 3D-HEVC[1] [3]. The dependency between the encoded views, due to the use of interview prediction, makes such a standard impractical for IMVS applications. Because more than one view, are in general, needed to decode a single view, thereby increasing the network traffic requirements. Depth-image-based rendering (DIBR) [4] offers a promising solution, where additional non-captured viewpoints could be synthesized using captured texture and depth maps. This advantage, coupled with an efficient coding, motivates the use of Multi-view Video plus Depth (MVD) format in this paper [5].

In MVD-based IMVS systems, compressed depth maps are used together with the associated texture views to synthesize the virtual views. However, in delay-sensitive wireless video communication under high packet error probability (PEP), e.g., in a wireless fading channel, transmission distortion is caused by packet errors during the transmission of a video sequence. This means that if a transmission error affects depth and/or texture data, virtual views will suffer from this error.

In particular, the geometry distortion of depth maps could cause erroneous warping of texture pixels, which may lead to unexpected occlusions, holes or annoying artifacts in the rendered virtual view [6]. Hence, in unreliable underlying networks, channel errors severely deteriorate the quality experienced by the end user while switching views. How to improve the robustness of texture and depth images in IMVS applications over error-prone networks is an open challenge.

* Corresponding author.
E-mail address: yzhao@bjtu.edu.cn (Y. Zhao).

[1] In MV-HEVC and 3D-HEVC, one view is selected as the base view to be independently coded using HEVC. Other views are coded using inter-view prediction based on the base view or other reference views to effectively reduce redundant information between views.

https://doi.org/10.1016/j.sigpro.2017.12.016
0165-1684/

Extensive efforts have been dedicated to address the error resiliency problem of 3D videos for decades. However, most works focus on optimizing coding structure for robust transmission, or designing efficient schemes for error concealment. They directly encapsulate texture and depth data into separate packets and then deliver them over a lossy channel, discarding the DIBR technique characteristics. The key concept is that, if a part of the texture frame is lost, to render a virtual view, its counterpart in the depth map cannot be exploited, and vice versa is also true. Therefore, the texture data and the associated depth are supposed to be encapsulated in the same packet.

In this paper, given nonstationary random packet losses in a wireless fading channel, we address an error-resilient MVD-based video communication. Without retransmission, the system aims at the real-time streaming applications of multiview video content, such as live sports[2]. In more details, we consider a scenario where multiview data is streamed to the clients by the server (e.g., Twitch, Live TV). To make the system with lower complexity and high efficiency, we replace the unicast connections to every client with the wireless multicast delivery platform [7]. It means that a client subscribes to only one multicast channel that distributes his current desired view, and subscribes to a different channel for another view. To alleviate the burden on the server and to reduce the costs on cameras infrastructure, dense views are not considered and the intermediate views are synthesized from their neighbors. In this case, the user subscribes to the two adjacent views via different channels, and uses DIBR to synthesize the intermediate view-to-be-watched. When the user requests a virtual view or an actual view, what is the best packet encapsulation mechanism? It is the topic of this paper. A novel packetization strategy as an efficient solution is proposed, where texture data and the associated depth go into the same packet. To this end, both texture and depth images are firstly encoded by video codecs such as H.264/AVC [8] to collect the size of each macroblocks (MBs). Then we need to determine the MBs allocation for each packet while encoding, so as to ensure that the size of included Texture-MBs (T-MBs) plus associated Depth-MBs (D-MBs) is less than the maximum transfer unit (MTU). We employed the conventional approach where texture and depth data is encapsulated in separate packets as the comparison benchmark.

Different loss events have diverse effects on the resulting distortion, which is crucial for designing and analyzing video communication systems [9]. To understand the unequal effects between isolated losses and burst losses, we study the distortion of different packetization strategies in terms of not only the global loss case but also various loss patterns of common interest. Specifically, burst losses and losses divided by a certain lag are considered. In addition, the effects of burst losses are significantly diverse for videos encoded at different rates generally. For example, in the lower rate case, a full frame could fit within a single packet, and thereby burst losses can lead to multiple video frames to be lost. On the other hand, each frame might be coded into several packets in the higher rate case, resulting in only a partial loss in a single frame. Thus, we further analyze the resulting difference between two packetization schemes when encoding views into different rate levels, i.e., a single frame is encapsulated into one packet or up to multiple packets. Finally, the performances of joint and separate packetization strategy are compared for two packet loss model settings in IMVS wireless multicast, the Gilbert–Elliott (*GE*) model and independent and identically distributed (*iid*) model.

[2] Delay-sensitive wireless video communication usually does not allow to correct packet errors and alleviate error propagation by the retransmission, since retransmission may cause long delay.

In conclusion, the contributions of this paper are the following:

- To improve error-resilience capabilities, we propose a novel packetization strategy to increase the navigation quality experienced by users in MVD-based IMVS applications. In this packetization scheme, texture data and the corresponding depth go into the same packet.
- In order to clearly understand the real effects of different packetization strategies, we examine them under different scenario settings. The packet loss patterns of interest are firstly identified, and in particular, burst losses as important cases are inserted. Then different video bit rates are considered. Two packet loss models are finally employed in the wireless channel to characterize different probabilities for each loss event.
- Experimental results show that two packetization strategies have their merits in different cases. However, compared with the currently used approach for encapsulating texture and depth data in separate packets, the proposed joint packetization scheme can achieve up to 2.32 dB and 1.55 dB quality improvement for texture views and synthesized views, respectively.

To the best of our knowledge, we are the first to propose such a joint packetization strategy for texture and depth data as well as to thoroughly analyze the performances of different packetization schemes in both several cases of interest and general case. The navigation quality experienced by the user is examined, where both the texture view quality and synthesized view quality are included. This work aims to put forward guidelines for practical design of packetization strategy in IMVS systems or other multiview video streaming frameworks to enhance loss-resilient capabilities.

The rest of this paper is organized as follows. Related works are introduced in Section 2. Section 3 presents the IMVS system structure. In Section 4, we describe and analyze two packetization strategies for texture and depth data. The experimental results are shown and discussed in Section 5, while Section 6 draws the concluding remarks.

2. Related works

Studies to date have investigated work for interactive multiview video streaming applications from different perspectives. From the encoding side, merge frame using piecewise constant functions [10] and an efficient coding structure with replication optimization [11] are designed to facilitate view switching. The navigation segment representations [12] and adaptive streaming representations [13] are optimized for storage in the server from the representation optimization side. From the view selection side, in-network view synthesis [14], the positioning of the reference views [15] and the content characteristics as well as the client navigation properties [16] are considered in optimizing the receiver control. However, none of these works focuses on the packetization strategy before. Many works have been published on the packetization strategy before. In the following, several typical schemes are presented from the encoder and decoder side perspective.

From the encoder side perspective, the main open challenges on 3D video streaming over lossy network are: how to efficiently encode captured sources, how to efficiently synthesize virtual views, and how to robustly deliver information to users, such that error propagation can be halted. Recently, a number of rate-distortion optimized coding algorithms have been studied, where the estimated end-to-end distortion models are developed. For MVC techniques, a new coding tool by inserting a unified distributed source coding (uDSC) frame into the multiview coding structure is proposed in [17]. It facilitates view switching and loss-resilient capabilities in IMVS applications. Meanwhile, packets are scheduled for network transmission based on a rate-distortion model. Aiming at the robust multiview video transmission over