Computers and Electrical Engineering 72 (2018) 361–368

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Tile-based 360-degree video streaming for mobile virtual reality in cyber physical system☆

Jangwoo Son, Eun-Seok Ryu*

Department of Computer Engineering, Gachon University, 1342 Seongnamdaero, Sujeong-gu, Seongnam, Gyeonggi 13120, Republic of Korea

A R T I C L E I N F O

Article history:
Received 18 September 2017
Revised 30 June 2018
Accepted 1 October 2018

Keywords:
VR
Tile
region of interest (ROI)
motion-constrained tile set (MCTS)
SHVC

A B S T R A C T

Today, the demand for and interest in virtual reality (VR) is increasing, since we can now easily experience VR in many applications. However, the computational ability of mobile VR is limited compared to that of tethered VR. Since VR represents a 360-degree area, providing high quality only for the area viewed by the user saves considerable bandwidth. Therefore, we propose a new tile-based streaming method that transforms 360-degree videos into mobile VR using high efficiency video coding (HEVC) and the scalability extension of HEVC (SHVC). While the SHVC base layer (BL) represents the entire picture, the enhancement layer (EL) can transmit only the desired tiles by applying the proposed method. By transmitting the BL and EL using region of interest (ROI) tiles, the proposed method helps reduce not only the computational complexity on the decoder side but also the network bandwidth.

1. Introduction

In recent years, many companies have launched head-mounted displays (HMDs), and new standards are being created for 360-degree video streaming.

HMDs are display devices worn on the head and have a display optic in front of one or each eye. These devices support head tracking to provide a 360-degree view and therefore require high-quality and high-performance hardware. The recommended specification for Oculus Rift, a type of tethered virtual reality (VR) system, is the Intel i5-4590, Nvidia GeForce GTX 970 processor with 8 GB RAM. In contrast to a tethered VR system, which is a PC-based HMD, a mobile VR system such as Samsung's Gear VR headset has limited video processing capabilities. Table 1 illustrates the differences between mobile VR and tethered VR systems.

Table 1 shows that, while a mobile VR system is more convenient, its performance is poor compared to that of a tethered VR system. To increase the video processing efficiency with a limited specification, we propose a method to solve the problems of bitrate and computational complexity through region of interest (ROI)-based SHVC tile processing. We propose a solution to the motion compensation problem that occurs when the enhancement layer (EL) sends selected ROI tiles and the base layer (BL) sends a full picture using an SHVC encoder. Furthermore, we propose a method of sending selected ROI tiles in a single layer using the HEVC encoder.

☆ Reviews processed and recommended for publication to the Editor–in–Chief by Guest Editor Dr. Jia–Shing Sheu.

* Corresponding author.
E-mail address: esryu@gachon.ac.kr (E.-S. Ryu).

https://doi.org/10.1016/j.compeleceng.2018.10.002

Table 1
Differences between mobile VR and tethered VR systems.

	Mobile VR	Tethered VR
Pros	Wireless	Computing power
	Portability	Various content
Cons	Less-capable tracking	Limited portability
	Performance ∝smartphone	Expensive

The composition of the paper is as follows. Section 2 gives a brief description of 360 video standards and ROI-related research. Section 3 describes the architecture of the proposed methods. Section 4 describes the implementation process, and Section 5 shows the performance of each technology.

2. Related work

2.1. 360 video standards of MPEG, JCT-VC, and JVET

The Moving Picture Experts Group (MPEG), the Joint Collaborative Team on Video Coding (JCT-VC), and the Joint Video Exploration Team (JVET) have discussed various 360-degree video streaming standards for VR. JVET defines Common Test Conditions (CTC) and evaluation procedures for 360 video [1]. Since VR requires high quality resolution, the test sequence is composed of 4 K and 8 K. In addition, MPEG-I (MPEG Immersive media) introduced the three-step goals for 360 video [2]. In the first phase, the aim of MPEG-I was to complete a 3 Degree of Freedom (3DoF) standard by 2017. In the second phase, their goal is to activate VR commercial services and to support 3DoF+by 2020. The objective of the last phase is to support 6DoF by 2022. This allows the user's movements to be reflected in VR. In addition, MPEG DASH-VR standardized the dynamic adaptive streaming over http (DASH) syntax for VR. They configured five used cases for compatibility and efficient streaming, one of which is viewport-based DASH streaming for VR content [3]. In addition to DASH, the viewport users observe is one of the key points according to VR standards for reducing bandwidth. To this end, the standardization groups have discussed the possibility of motion-constrained tile sets (MCTS) [4].

2.2. Single encoding based on MCTS

Unlike the conventional encoder, the MCTS-applied encoder does not temporally refer to tiles that have different positions on the current picture and the reference picture. Thus, the tiles can be separated in one bitstream, although the bitrate increases slightly. A. Zare et al. explains a method of saving bitrate when sending only the field of view (FOV) area using the MCTS-applied encoder [5]. In their study, the MCTS method is applied and the encoding efficiency is reduced by from 3% to 6%. However, the study reduces the bitrate by from 30% to 40% when transmitting tiles corresponding to FOV. Compared with their study, our study embodies the installation process for applying MCTS to the HEVC reference software (HM) and SHVC reference software (SHM), and describes implementation issues.

2.3. Tile based panoramic streaming using SHVC

Y. Sanchez et al. proposed a technique to minimize picture transition delay and bitrate according to the change of ROI, which is a point seen by the user when using SHVC [6]. Their technique involved dividing BL and EL into multiple tiles, and only the tiles corresponding to ROI are streamed. However, if streaming only the corresponding tiles, a prediction mismatch occurs when decoding. Fig. 1 depicts the prediction mismatch and its solution. At the encoder, the second tile of the t1 picture refers to the second tile of the t0 picture. Considering the ROI, the t0 picture transmits from the second to the fourth tiles, and the t1 picture transmits from the first to the third tiles. The decoder encounters prediction mismatch with reference to the same second tile using the encoder's motion vector. This study creates a Generated Reference Picture (GRP) between the reference and the current picture in order to correct the motion vector. The GRP holds motion vector information that compensates for the prediction mismatch that occurs when decoding only some of the tiles. The Multi Layer GRP (MLGRP) is similar to GRP, but utilizes the characteristics of SHVC to obtain motion vector information through the lower layer. This study solves the problem of motion vectors, but there is an overhead of generating GRP. In contrast, we solve the motion vector problem in the encoder and perform a single encoding.

2.4. Viewport independent studies on 360-degree video

The viewport independent methods transmit whole 360-degree video such as equirectangular projection (ERP) and cubemap projection (CMP). These methods reduce bitrates and computational complexity by down-sampling and/or increasing the number of quantization parameters (QPs) of lesser important regions. Fig. 2 shows the efficient preprocessing studies using the ERP and CMP regions. The adaptive-QP ERP is encoded into different qualities for each region in consideration of the user's gaze. The region-wise packing also considers the user's gaze. The ERP down-samples the top and bottom regions,