Signal Processing: Image Communication 57 (2017) 157–172

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

QoE-driven optimization for cloud-assisted DASH-based scalable interactive multiview video streaming over wireless network

Mincheng Zhao [a,c,*], Xiangyang Gong [a], Jie Liang [b], Wendong Wang [a], Xirong Que [a], Yihua Guo [a], Shiduan Cheng [a]

[a] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China
[b] School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada
[c] The 20th Research Institute of CETC, Xi'an, 710068, China

ARTICLE INFO

Keywords:
Dynamic adaptive streaming over HTTP (DASH)
Interactive multiview video streaming (IMVS)
Quality of experience (QoE)
Cross-layer design
Resource allocation

ABSTRACT

In interactive multiview video streaming (IMVS), the viewers can periodically switch viewpoints. If the captured view is not available at the desired viewpoint, virtual views can be rendered from neighboring coded views using view synthesis techniques. Dynamic adaptive streaming over HTTP (DASH) is a new standard that allows to adjust the quality of video streaming based on the network condition. In this paper, an improved DASH-based IMVS framework is proposed. It has the following characteristics. First, virtual views could be adaptively generated at either the cloud-based server or the client in our scheme, depending on the network condition and the cost of the cloud. Second, scalable video coding (SVC) is used to improve the flexibility in our system. To optimize the Quality of Experience (QoE) of multiple clients in wireless scenario, we develop a cross-layer optimization scheme. We first propose a new cache management method to selectively store the video data according to SVC structure and the clients' requesting condition. Next, a cross-layer scheduling scheme is proposed by considering the video rate adaptation and the wireless resource allocation. The optimization problem is shown to be equivalent to the Multiple Choice Knapsack (MCKP) problem. A dynamic programming method and a low-cost greedy method are developed to solve the problem. Simulations with the NS3 tool demonstrate the advantage of our proposed scheme over the existing approach that always uses client-based view synthesis and single-layer video coding.

1. Introduction

Multiview video has attracted significant attention, where the scene is captured by multiple cameras from different viewpoints. Interactive Multiview Video Streaming (IMVS) or Free Viewpoint Video (FVV) streaming [1 –3] is one interesting application of multiview videos, it allows the viewer to switch the viewpoint during playback to get motion parallax. In order to reduce the storage cost and still provide smooth view switching, it can only store some selected views at the server. When the requested view is not directly available at the server, it is synthesized using the two neighboring views. The synthesized view is called virtual view, which can be generated using the two neighboring texture views via disparity-based methods, or using both the texture and depth videos of the two neighboring views via depth-image-based rendering (DIBR) techniques [4]. The latter approach generally has better quality, and is therefore adopted in this paper. On the other hand, view synthesis can be performed at the server or the client. In the latter case, both the texture and depth videos of the neighboring views should be encoded and transmitted to the client.

The dynamic adaptive streaming over HTTP (DASH) [5] is a new international standard for video delivery. In DASH, each video is divided into segments, and each segment is encoded into multiple bit streams or quality levels. Detailed information of all segments is described in a XML Media Presentation Description (MPD) file. When the DASH starts, the MPD file is firstly downloaded to the client. Then the client adaptively chooses the appropriate version of a segment to download according to network conditions, *e.g.*, instantaneous TCP throughput. The client adaptation policy should meet various goals, such as avoiding playback interruption, reducing frequent changes of quality levels, and

* Corresponding author at: State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China.
E-mail addresses: mczhao@bupt.edu.cn (M. Zhao), xygong@bupt.edu.cn (X. Gong), jiel@sfu.ca (J. Liang), wdwang@bupt.edu.cn (W. Wang), rongqx@bupt.edu.cn (X. Que), guoyh@bupt.edu.cn (Y. Guo), chsd@bupt.edu.cn (S. Cheng).

http://dx.doi.org/10.1016/j.image.2017.05.015
Received 6 December 2016; Received in revised form 24 May 2017; Accepted 28 May 2017
0923-5965/

maximizing the average video quality. When multiple DASH clients compete for the shared bandwidth, the system also needs to address the fairness and bandwidth under-utilization issues. In wireless networks, it also should consider the wireless rate allocation to improve the DASH transmission performance, because if the allocated rate is not reasonable, it is hard to improve the total transmission quality by only modifying the client adaptation method. DASH transcoding management is another important issue. It has the following targets: (1) reducing the transcoding latency by setting up reasonable cloud resource scheduling method in the cloud. (2) reducing the storage cost. It can be realized by only saving the most popular versions of a video, and other versions will be transcoded in real time according to the users' requests.

In [6], a DASH-based interactive free viewpoint streaming system was proposed. Its rate adaptation scheme aims to maximize the quality of the virtual view, based on the empirical bitrate-distortion curve that represents the distortion of the virtual view according to the bit rates of the texture and depth maps of the reference views. However, only non-scalable video coding is used in [6], and the view synthesis is always performed at the client, which requires the transmission of both texture and depth videos of two reference views; hence the bandwidth requirement is higher than the single-view video streaming. It also has higher requirement on the hardware capability of the client.

In this work, we propose a cloud-assisted and DASH-based Scalable Interactive Multiview Video Streaming (CDS-IMVS) system. Compared to previous DASH-based single-view video transmission, it needs to consider the virtual view synthesis and the related problems. The system adopts scalable video coding (SVC) to facilitate DASH transmission. The texture and depth videos of all captured views are coded using H.264 SVC and saved in layered structure. The advantage of SVC in DASH has been demonstrated in [7], using SVC in DASH-based IMVS can improve the flexibility and avoid playback interruption effectively. The details are discussed in Section 3.3. Moreover, different from [6], where the virtual view is always generated at the client, we allow the synthesized view to be generated at either the client or the server, depending on the client's capability and the server cost. When the synthesized view is generated at the server, only the virtual view is coded and transmitted. This reduces the bandwidth requirement, allows the system to support more users, and reduces the possible playback interruptions when the network bandwidth is limited (which is important in wireless scenario). It also reduces the client power consumption (which is important for mobile users), and provides virtual view switching capability for low-end clients that cannot afford real-time view synthesis. However, the server-synthesized virtual-view video also needs to be transcoded according to DASH layer structure before transmitted, which will also increase the cloud operating cost. In order to resolve the problem, we design an adaptive view synthesis (AVS) method for the CDS-IMVS system, which makes the decision based on the estimated available network bandwidth and the cost of virtual view synthesis in the cloud.

In wireless network, AVS method could improve one user's QoE and reduces the bandwidth requirement based on the user's allocated rate. However, it could not get the best wireless system performance without integrating media delivery and wireless radio-level adaptation. In order to improve the QoE for multiple users in wireless transmission scenario, this paper also develops a cross-layer scheduling scheme by jointly considering the bitrate adaptation, wireless access rate allocation, DASH transcoding cost, cloud bandwidth and storage costs. First, in order to reduce the computing and storage cost, a new cache management method is developed, which selectively stores the video data based on the SVC structure, the condition of clients' request, the storage cost and the DASH transcoding cost. Second, a cross-layer bitrate scheduler method (CBS) for CDS-IMVS system is proposed to select the bitrate level for multiple wireless clients. Moreover, a MAC-layer allocator is used to allocate wireless rate to ensure that the allocated wireless access rate can achieve the calculated bitrate by the CBS method. The CBS problem is shown to be equivalent to the Multiple Choice Knapsack (MCKP) problem, which can be solved by a dynamic programming method. In order to reduce the computing complexity, a greedy method is also proposed to solve the problem. Simulations with the NS3 tool demonstrate the advantage of our proposed scheme over the existing approach in [6].

Some preliminary results of this paper are reported in [8], mainly on the adaptive view synthesis part. In this paper, we develop more accurate cloud transcoding model and cloud bandwidth and storage cost models, add the entire Section 5, and include more comprehensive experimental results.

The paper is organized as follows. We firstly overview the related technologies in Section 2. Section 3 introduces the proposed CDS-IMVS framework. Section 4 describes the proposed adaptive virtual view synthesis method. Section 5 introduces the cross-layer design framework, including the cache management, the bitrate scheduling, two fast solutions, and the resource allocation. Simulation results are presented in Section 6, followed by concluding remarks in Section 7.

2. Related technologies

2.1. Interactive multiview video streaming

In multiview videos [9], multiple captured views are compressed by taking advantage of the redundancies among the different views. To reduce the transmission rate, IMVS does not transmit all the captured viewpoints. IMVS systems have the following challenges [6]: view switching latency, scalability, bandwidth variation and immersive experience. The corresponding optimization becomes more challenging, as we have more factors to consider. In [10], coarse and fine quality layers of several views are grouped and pre-encoded. It will estimate the users behavior first, and then transmit a subset of views of low quality plus two views of high quality to the client. All transmitted views were subsequently decoded, and the high quality views requested by the client are displayed first, and other low-quality views will be displayed if the client decides to switch to them. Xiu et al. [3] propose a joint optimization of the frame coding structure, transmission schedule, and quantization parameters of the texture and depth maps of multiple camera views for IMVS system. It could realize zero-delay view switching and reduce effectively the required transmission rate.

In order to reduce the transmission bitrate for multiple users due to the redundant transmission of overlapping frames, the authors in [11] proposed User dependent Multi-view video Streaming for Multi-users (UMSM). The overlapping frames required by multiple users are transmitted by multicast in the UMSM method. However, the clients have to watch the videos at the same point-in-time for the multicast transmission. The client could watch the video at any point-in-time using unicast. The authors in [12] and [13] discuss the DASH video and FVV transmission over LTE network respectively. In order to reduce the bandwidth requirement, they propose to use a hybrid unicast and multicast method based on the wireless condition and the clients' requests.

2.2. Dynamic adaptive streaming over HTTP

Recently, some authors began to investigate the DASH QoE [14,15] and the QoE-driven DASH optimization [16]. The segment duration and the client adaptation strategy in DASH greatly influence the user's QoE [14,17]. Several client adaptation methods were proposed in [18–20], where non-scalable video coding is used in [18,19], and SVC is used in [20]. With SVC, a video can be encoded once, but decoded in different ways, based on factors such as network condition and the decoder's capability. Using SVC in DASH allows the client to download enhancement layers of an already received segment when upgrading to a higher quality, thereby improving the flexibility and reducing the chance of wrong decision in selecting an appropriate version of a segment, especially in dynamic wireless conditions. In [18], the bitrate