OmniViewer: Enabling Multi-modal 3D DASH

Zhenhuan Gao Shannon Chen Klara Nahrstedt

Department of Computer Science
University of Illinois at Urbana-Champaign
{zgao11, cchen116, klara}@illinois.edu

ABSTRACT

This paper presents OmniViewer, a multi-modal 3D video streaming system based on Dynamic Adaptive Streaming over HTTP (DASH) standard. OmniViewer allows users to view arbitrary side of a performer by choosing the view angle from 0°to 360°. Besides, according to the current available bandwidth, it can also adaptively change the bitrate of rendered 3D video for both smooth and high-quality view rendering. Finally, OmniViewer extends traditional DASH implementation to support multi-modal data streaming besides video and audio.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Artificial, augmented, and virtual realities

Keywords

3D; Real-time; DASH; View

1. INTRODUCTION

3D Tele-immersion (3DTI) system enables remote users to view the local scene and even to interact with local users in a virtual 3D space, which has promoted lots of potential applications such as remote physiotherapy [1]. Furthermore, in a traditional 2D on-demand video streaming system, the interactivity of the end users choosing any viewpoint is not available except if the number of streams is equal to the number of possible views. In contrast, 3DTI enables interactivity of users from any viewpoint. Thus, free view is the trump of 3D over 2D. Gao et al. presented FreeViewer [2], a 3DTI system offering free-view feature. However, transmission of such huge 3D data without encoding necessitates high bandwidth and is not feasible for commercial use. Therefore, 3D video recording, compresssion and streaming are further challenges in 3D systems.

To the best of our knowledge, there is no 3D video streaming standard which can fit the rendering and streaming part of 3D systems like 3DTI system. We also notice that in 2D video streaming, the popularity of DASH (defined by ISO/IEC 23009-1) skyrockets and an increasing number of commercial streaming services are adopting this new standard. While the benefits of DASH is clear for 2D streaming, the research of DASH in 3DTI is not well understood. Therefore, we present OmniViewer, a complete multi-modal on-demand 3D video streaming solution based on DASH.

Same as with FreeViewer, OmniViewer uses four 3D camera streams to cover the whole 360°view. Each 3D stream is encoded with multiple bitrates as a normal 2D stream, containing a pair of color frame and depth frame. The user is able to set any view angle to watch the performer from. Also, to be consistent with DASH standard, Omniviewer can adaptively change the bitrates of streams to achieve the best possible media quality of the chosen view. Moreover, since fundamentally DASH is ignorant of content and encoder/decoder, we introduce multi-modal data (eg. Kinect skeleton, heartbeat sensor data) into OmniViewer to support multi-modal media streaming, expanding the scope of DASH to more general use cases.

In summary, the contributions of Omniviewer lie in four major aspects:

- 3D video recording. OmniViewer leverages 2D video en-/decoding technique to store and retrieve monoscopic 3D video segments in a natural way.
- Omni-view rendering. OmniViewer offers the user to view the 3D video from any view angle at any time through a natural user interface.
- Adaptive 3D streaming. On top of DASH, OmniViewer always provisions the best media quality while keeps the total bitrate under current available bandwidth.
- Multi-modal media support. OmniViewer supports any type of media including but not limiting to video and audio giving more flexibility in applications.

2. SYSTEM OVERVIEW

This section gives an overview of the system architecture including both server and client sides.

2.1 OmniViewer Server

2.1.1 Recording

As illustrated in Figure 1(a), we still adopt four 3D cameras (Kinect). In order to fit DASH, we represent each 3D camera video as a 2D video where each 2D frame consists of

MM'15, October 26–30, 2015, Brisbane, Australia.
ACM 978-1-4503-3459-4/15/10.
DOI: http://dx.doi.org/10.1145/2733373.2807971 .

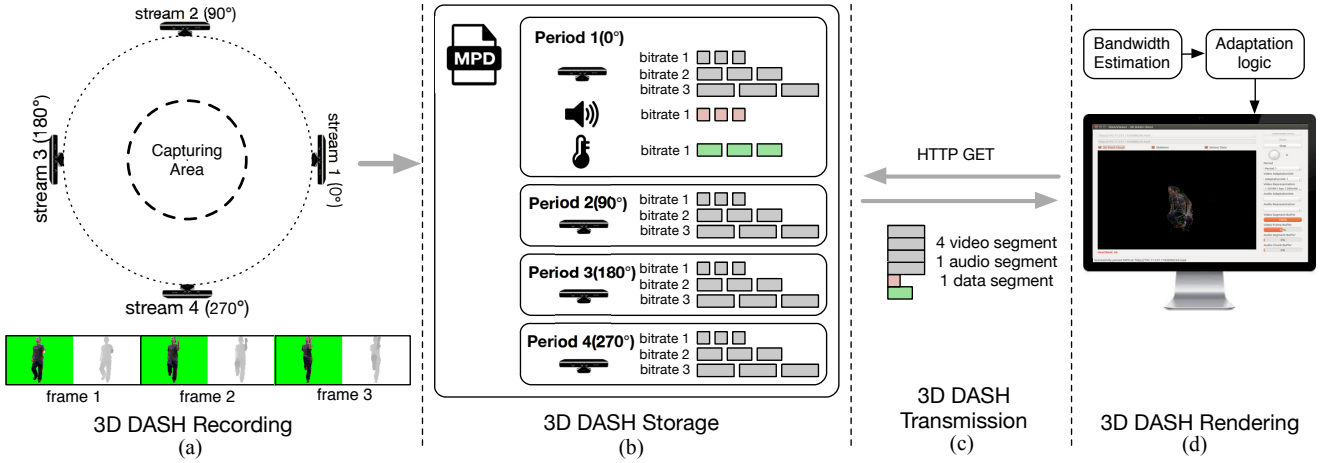

Figure 1: OmniViewer System Overview

a pair of color (RGB) and depth (Gray) frames, stitched together horizontally. Then we compress these four 2D videos using H.264 codec with different bitrates and slice them into 1s DASH segments using MP4Box [3]. For audio, we only need one stream and apply similar operation. For other media data, in our case we incorporate skeleton and heartbeat data, encoding them in JSON file, one file per segment including the data for each frame.

2.1.2 Storage

In DASH, a media presentation description (MPD) file is required to describe segment information. To be specific, the media presentation defines the video sequence with one or more consecutive periods that break up the video from start to finish. Each period contains one or more adaptation sets that contain the content of one modality data, video, audio, etc. Each adaptation set contains multiple representations, which represents the streams with different qualities.

In 3D DASH, since we have four video streams, OmniViewer leverages period customization to support streaming four video streams concurrently. We configure the MPD file in OmniViewer system with four periods and each of them contains a video adaptation set consisting of the same number of representations with the same qualities. For example, in Figure 1(b), we show a MPD for a 3D video that includes 4 camera-related periods, and each segment in each period includes data of 1s under different bitrates (the small blocks). As for audio and other data, we only need one stream for each modality. Thus, we choose to put their adaptation sets into the first period (period 1(0°) in Figure 1(b)) and let the client read them from the first period. Considering DASH is audio/video codec agnostic, we set other data segment in the form of JSON and implement our own en-/decoder in the recording step and in the client. All the generated segments and customized MPD file are stored in a standard static webserver.

2.2 OmniViewer Client

We implement all the components of OmniViewer client based on the open-source sample player that comes with libdash [4], an open-source C++ library which implements the full MPEG-DASH standard defined by ISO/IEC 23009-1. After the OmniViewer client downloads the MPD file, it analyzes the information of periods, adaptation sets and representations. Then, it opens two decoders, audio/video decoder and our own JSON decoder, and six buffers, four for videos, another for audio, and the last one for other data except for A/V data. When all the data, required for the current frame, are available, the frame will be rendered in an OpenGL 3D context. At any time, the human viewer is able to choose the preferred view angle from the OmniViewer client GUI. Then the OmniViewer client will select the required two neighbouring video streams, decode the 3D point clouds, merge them into a partial 3D model by geometrical relationship and finally render the frame according to the computed viewpoint. Also, when the bandwidth fluctuates greatly, the client will correspondingly adapt the bitrates of media streams to keep smooth and high-quality rendering.

3. ACKNOWLEDGEMENTS

This presented work does not represent opinions of NSF. The presented material is based on our work supported by NSF Grant CNS 10-12194, CNS 09-64081 and Deutsche Telecom Gift funding.

4. CONCLUSIONS

In this demo paper, we present OmniViewer, the multi-modal 3D DASH framwork, that supports free-view, adaptive, multi-modal 3D video streaming. OmniViewer combines DASH, 3D and multi-modal data together, which expands the current scope of both DASH and 3D systems. The versatility and flexibility of OmniViewer can be leveraged to fit many use cases, such as remote physiotherapy, entertainmnet and other 3D activities.

5. REFERENCES

[1] Klara Nahrstedt. 3D Teleimmersion for Remote Injury Assessment. In *Proceedings of the 2012 International Workshop on Socially-aware Multimedia*, SAM '12, pages 21–24, New York, NY, USA, 2012. ACM.

[2] Zhenhuan Gao, Shannon Chen, and Klara Nahrstedt. FreeViewer: An Intelligent Director for 3D Tele-Immersion System. In *Proceedings of the ACM International Conference on Multimedia*, MM '14, pages 755–756, New York, NY, USA, 2014. ACM.

[3] MP4Box. `http://gpac.wp.mines-telecom.fr/mp4box/`.

[4] Libdash. `http://github.com/bitmovin/libdash`.