Free-viewpoint Relationship Description based Streaming Systems for Arbitrary View Switching

Seulki Song, Yonghwan Kim, Yong-Suk Park, and Jungwook Wee
Korea Electronics Technology Institute
Seoul, Korea
Email: {seulkiemma, yonghwan, yspark, jwwee}@keti.re.kr

***Abstract*—Recent advances in video processing and transmitting technologies have increased user interest and demand for realistic services. Among realistic services, the free-viewpoint video (FVV) streaming service has been studied. In order to provide the FVV streaming service, it is necessary to distinguish the number of videos and provide adjacent views for view switching. In this paper, we design and propose a structure that enables the client to know the relationship between views in free-viewpoint relationship description (FRD) architecture. The proposed FRD is used to implement the FVV streaming system, and the possibility of rapid new switching is verified.**

***Index Terms*—Free-viewpoint video, MPD, DASH, Streaming system**

I. Introduction

Recently, advancements in video processing and transmitting technologies have increased user interest and demand for realistic media services. The realistic media services such as VR/AR, 3D video, and free-viewpoint video (FVV), which allow users to feel immersion and realism. Since the FVV allows user to select direction, viewpoint, and position freely, it can provide the image and sound of the position desired by the user. In addition, it is possible to move the various viewpoints by synthesizing an intermediate viewpoint image using two adjacent camera images [1]. To provide the FVV service, a streaming technology between a contents server and a client is required. This can be implemented using an dynamic adaptive streaming over HTTP (DASH) standard [2], [3]. The DASH is a technology that divides contents with various resolutions into segments of a certain length and transmits images with resolution suitable for the network capacity. When applied to the FVV service, it provides an easy way to switch contents by moving viewpoint. Since the traditional DASH technology is not appropriate for the FVV streaming service, a free-viewpoint relationship description (FRD) for distinguishing and transmitting multiple viewpoint images has been proposed [4]. The FRD defines *View* and *Sub-view* to distinguish each image of the FVV, and defines ID of each *Sub-view* and position of image. In [4], the position of image is represented by the absolute coordinates at which the image is acquired. However, it is difficult to find out the association information of adjacent view due to the lack of a relevant information in the client. Therefore, it is necessary to provide additional information that can easily obtain the adjacent image when the client moves the viewpoint.

In this paper, we propose the structure that provides information on the relationship between views using a relative coordinate in the FRD structure. The proposed structure includes absolute coordinate based camera position information for synthesizing intermediate image and relative coordinate for streaming the adjacent image.

This paper is organized as follows. The free-viewpoint video streaming system are briefly reviewed in Section II. In Section III, the proposed structure is described. Experimental results of the proposed structure are presented in Section IV. Section V concludes this paper.

This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2018 [R2017030041, Experience Maximization Technology for Cultural Contents at Free Selection View]

II. Free-viewpoint Video Streaming System

The FVV streaming service provides amusement to the users by moving viewpoint. The user expects to navigate freely and continuously to any viewpoint. To provide this, when taking video, cameras are placed at regular intervals to obtain images from various viewpoints. The non-captured image is generated by synthesizing intermediate image using two adjacent camera images. If intermediate images are provided between the images taken, the user moves smooth their viewpoint.

It is not necessary to acquire or compose an image at all viewpoints in order to provide the FVV streaming service. There are limits on server storage and transfer capacity to create images at any viewpoint. In addition, since the user does not simultaneously play all viewpoints, only images that an adjacent to the user's current viewpoint are necessary.

The images are acquired at maximum configurable intervals to provide the FVV streaming service. The synthesis of images is carried out in real time at the request of the user. The composition of the images can be done at the server side or at the client side [5]. In the first case, when the user's viewpoint moves, the images can be synthesized in real time at the server. The amount of network traffic is small, but there is a delay due to network transmission. In the second case, the images and composite information adjacent to the images being played by the client shall be streamed simultaneously. In addition, real-time synthesis should be performed to ensure natural view switching. It can be provided without delay since

978-1-5386-4646-5/18/$31.00 ©2018 IEEE

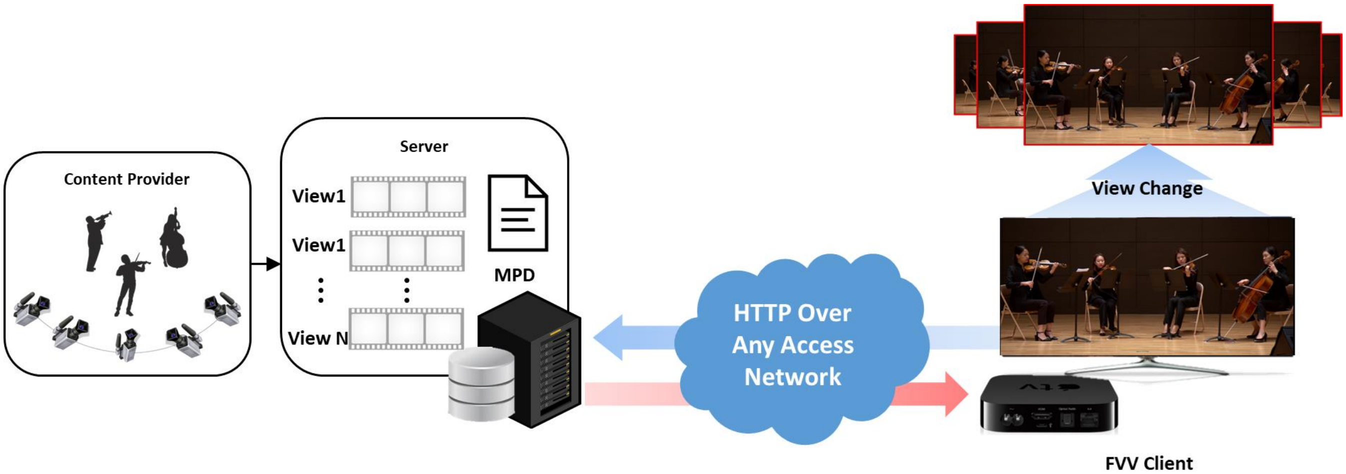

Fig. 1. The structure of the proposed free-viewpoint video streaming system.

several images can be streamed simultaneously and composed in real time.

In this paper, we design an FVV streaming system that the users can naturally move the various viewpoints by real-time synthesis at the client. The FVV streaming system with DASH is configured as shown in Fig. 1. Content providers acquire videos from various angles. The videos are generated into segments according to DASH and segment information are defined in MPD. The segments and MPD are stored on the server. The client receives initial view and adjacent views for view-synthesis with the start of the FVV service. Virtual viewpoint is generated via view-synthetic between adjacent views to provide natural viewpoint movements. When the user moves the viewpoint, the direction and the speed at which they move are measured. The client uses the measurements to predict and downloads the next viewpoint video [6].

III. Extended FRD Structure

The MPD is the actual medium for sending and receiving from servers and clients in DASH. The MPD documents provide information on media contents as well as segment components for clients to request when downloading each segment.

The MPD document includes elements such as *MPD*, *period*, *adaptation set*, *representation*, and *segment*. *MPD* consists of a number of *periods*, one *period* divided by an *adaptation set* according to the type of data, such as video and audio. Each *adaptation set* specifies the bit rate, resolution, and frame rate for the image and includes *representation*. *Representation* is generated by one or more depending on the transfer rate. This includes each segment address. The extended FRD (E-FRD) is created in the lower part of the adaptation set. The components contained in the FRD architecture match the representation.

The E-FRD includes information about adjacent view of client screen in the FVV streaming system. The proposed structure is described in the table I. The *ReferenceProperty* element refers to the view that serves as the reference for several neighboring views. It includes all of the characteristics of the type of video, number of cameras, and number of dimensions. It also contains the properties *siteValue*, which replaces the stage or stadium with coordinates to show the absolute position, and *relativeViewSite* which indicates relative position on the client screen shown in Fig. 2. In this figure, the coordinate values denote the two dimensional relative values, e.g. (left, right, up, down). The depth values for three dimensions and camera parameters are included. Fig. 3. shows an example of the E-FRD structure in MPD.

TABLE I
Semantics of FRD element

Element or Attribute Name	Description
ReferenceProperty	
@id	Specifies the id of the FRD. This shall be unique within the same viewpoint
@type	Parallel, curve, halfsphere
@total	Number of cameras
@siteValue	Camera coordinates of the stage/stadium
@dimension	Dimension of camera coordinates
@relativeViewSite	Relative position of screen
AdditionalProperty	
@id	Specifies the id of the FRD. This shall be unique within the same viewpoint
@siteValue	Site value of camera
@relativeViewSite	Relative position of screen

IV. Implementation Results

We implemented the components of the proposed FVV streaming system based on DASH. The FVV streaming system consists of DASH-based media server with the proposed MPD and a client that enables viewpoint movement of the FVV. Videos are encoded in 1920x1080 resolution, 30 fps, 3 Mbps bit rate and 30 GoP with HEVC. The encoded videos are segmented in 4 second lengths and stored with the MPD on the media server. The client acquires information on the

(1,0,1,0)	(0,0,1,0)	(0,1,1,0)
(1,0,0,0)	Main View (0,0,0,0)	(0,1,0,0)
(1,0,0,1)	(0,0,0,1)	(0,1,0,1)

Fig. 2. The view placement with coordinates.

```
<AdaptationSet group="1" maxBandwidth="2466086">
  <Role value="description">Inward View</Role>
  <SegmentTemplate startNumber="1" …/>
  <Representation id="1" mimeType="video/mp4" …/>
  <AdditionalProperty id="1" relativeViewSite="2,0,0,0"/>
  <AdditionalProperty id="2" relativeViewSite="1,0,0,0"/>
  <ReferenceProperty id="3" type="curve" dimension="2"
   total="5" relativeViewSite="0,0,0,0" />
  <AdditionalProperty id="4" relativeViewSite="0,1,0,0"/>
  <AdditionalProperty id="5" relativeViewSite="0,2,0,0"/>
</AdaptationSet>
<AdaptationSet group="2" maxBandwidth="2769246">
…
</AdaptationSet>
```

Fig. 3. The example of the FRD structure.

TABLE II
THE RESULT OF THE HEVC DECODING SPEED

Sequence	Bitrate [kbps]	speed [fps]
BasketballDrive	3936.06	239.83
Cacuts	3900.91	252.02
Kimono	3747.98	257.79
ParkScene	3816.03	211.64
Tennis	3771.73	280.37

FVV contents after receiving the MPD over HTTP. The client receives a simultaneous streaming of the main view and the adjacent views based on the acquired information. A fully optimized proprietary HEVC decoder is used, and the measurement results of the decoding speed in HEVC videos are shown in Table II. The decoding speed is the rate at which no screen playback was performed. As a result, the average speed with 4 threads is 248.33 fps.

Fig. 4 shows the download rate per segment of 3 videos from the client. Although the speed differs depending on the GoP unit and the characteristics of the video, the average speeds of the videos are not significantly different, as shown in Table III.

When moving the viewpoint, it takes 150ms from download

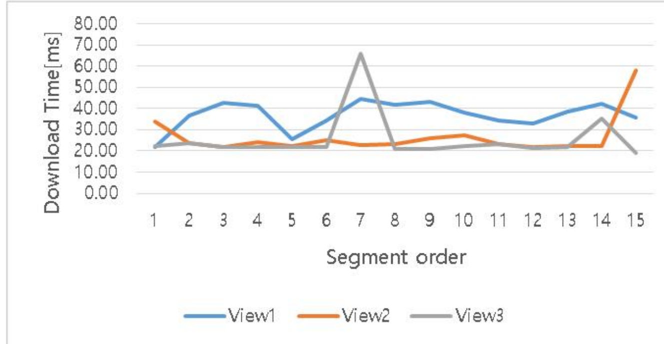

Fig. 4. The graph of the segment download time.

TABLE III
THE AVERAGE OF THE SEGMENT DOWNLOAD TIME

View	Average [ms]
View1	36.97
View2	26.62
View3	25.64

time to decode time, in the worst case scenario. Thus, when predicting the users time of movement, the user can continue to move viewpoints when videos are streamed within 0.2 seconds of the expected time of movement.

V. CONCLUSION

In this paper, the MPD architecture is designed and proposed to provide information on the relationship between images to enable the natural timing of the FVV streaming system. The FVV streaming service provides the optimal image and sound desired by the user. In order to provide the FVV streaming service, it is necessary to distinguish a number of images and provide adjacent images for client viewpoint movements. To do this, the architecture is designed to inform the relationship between images in the MPD architecture of DASH, and an FVV streaming system is implemented to confirm rapid viewpoint movements. In the future, a system that enables natural view switching, including the mid-image synthesis technique, will be implemented and tested.

REFERENCES

[1] A. Hamza and M. Hefeeda, "Adaptive Streaming of Interactive Free Viewpoint Videos to Heterogeneous Clients," *in Proc. 7th ACM Multimedia Syst. Conf*, May 2016, paper no. 10.

[2] Information technology - Dynamic adaptive streaming over HTTP (DASH) - part 1 : Media presentation description and segment formats, 2016

[3] T. Nunome and H. Tani, "Multi-View Video and Audio Transmission with MPEG-DASH and Its QoE," *in Proc. APCC*, Oct. 2015, pp. 575-579

[4] S. Song, Y. Park and J. Wee, "Design of Free-viewpoint Relationship Description based on MPEG-DASH," *in Proc. KSII ICONI*, Dec. 2017, pp. 391-392

[5] D. Yun and K. Chung, "DASH-based Multi-view Video Streaming System," *IEEE Trans. on Circuits and Systems for Video Technology*, pp. 1-6, Apr. 2017.

[6] A. Hamza and M. Hefeeda, "A DASH-based Free Viewpoint Video Streaming System," *in Proc. NOSSDAV*, Mar. 2014, pp. 1-6