STEREOSCOPIC SERVICE OF MULTIVIEW VIDEO FOR MOBILE ENVIRONMENTS

Junsup Kwon, Manbae Kim, and Changyeol Choi

Dept. of Computer Engineering, School of Information Technology
Kangwon National University
Chunchon, Korea
E-mail: manbae@kangwon.ac.kr

ABSTRACT

Multi-view video broadcasting have been conducted with a variety of different system architectures. However, the previous systems have focused on wired broadcasting. On the contrary, our system is designed to be applied to mobile 3DTV broadcasting. We present a service framework adaptable to mobile clients where the load on the client is reduced as much as possible. MPEG-21 multiview DIA description schemes are presented for the efficient adaptation between the serer and client. Furthermore, we examine the feasibility of the proposed framework by analyzing the frame rate that a client device can process.

Keywords: multiview video, MPEG-21 DIA, 3DTV

1. INTRODUCTION

Recently, as digital broadcasting technologies have rapidly progressed, users' expectations for more realistic and interactive broadcasting services have also increased [1, 2, 3]. Three dimensional televisions (3DTV) broadcasting is one of the promising technologies that could satisfy such expectations. 3D video service includes stereoscopic TV, 3D virtual studio, multi-view broadcasting and so forth. Among them, the multi-view broadcasting is a newly emerging area that can provide a variety of media services Simultaneously, the mobile devices have been spread rapidly. The service of the multiview video over mobile devices will provide a new paradigm over conventional services.

The main bottleneck is low channel bandwidth and processing capabilities of mobile devices. Therefore, an efficient framework is needed to overcome such problems. One scheme is to allocate most of processing functions to the server. Client can simply receive information form the server and thus client's overhead can be reduced.

This paper presents the service framework of multiview video over mobile environments. Since the multiview video are acquired from a finite number of cameras (e.g., 8, 10, etc), the IVR (intermediate view reconstruction) is usually required. Furthermore, to provide more effective transmission and message exchange at a server-client framework, MPEG-21 DIA (Digital Item Adaptation) is utilized with new multiview video descriptors.

This paper is organized as follows. In Section 2, the overall system architecture is presented. In Section 3, the system is explained in more details. Section 4 presents experimental results, followed by conclusion in Section 5.

2. SERVICE FRAMEWORK OVERVIEW

2.1 System Overview

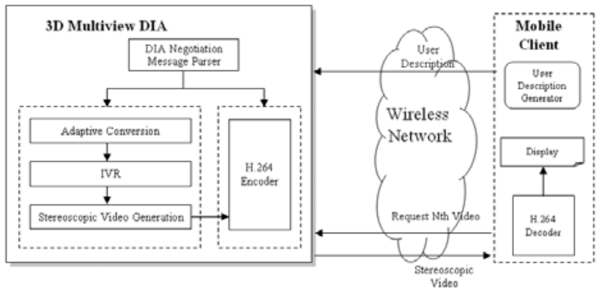

Fig. 1: The proposed service framework

The proposed service framework illustrated in Fig. 1 is composed of a 3D multiview video DIA, a server, and a mobile client.

The client has an H.264/AVC decoder and a user description generator. The 3D multiview video DIA implements the adaptive conversion such as intermediate view reconstruction, stereoscopic video generation and an H.264 encoding. The DIA negotiation message parser analyzes user descriptors in XML and requests each module to carry out its corresponding adaptation function.

When a mobile client requests *n*th view using MPEG-21 DIA DS (Description Scheme), the server adapts the video resolution and subsequently generates intermediate (virtual) views. Then it encodes the stereoscopic video by an H.264/AVC and delivers the compressed bitstream to the client.

2.2. MPEG-21 Multiview DIA

We propose efficient description tools for MPEG-21 DIA. These are composed of the descriptions for user preferences about multiview video presentation and for display capabilities. The proposed DIA description tools are categorized into two description schemes (DS) that functionally belong to Presentation Preferences and

Terminal Capabilities in Usage Environment Description Tools in MPEG-21 DIA [4, 5, 6]. In the Presentation Preferences, we propose effective descriptors related to the selection of a view number, a frame number, an intermediate view position, special effects, a parallax type, and a rendering format. In the Terminal Capabilities, the image resolution of a mobile device is included.

Fig. 2 depicts an XML schematic diagram of proposed DS, namely 3DMVDIA, under which two DSs are defined. *TerminalCapbilities* consists of two elements: *DisplayWidth* and *DisplayHeight*. If a client sends the information of a screen resolution, the DIA server adapts the input video according to the requested resolution.

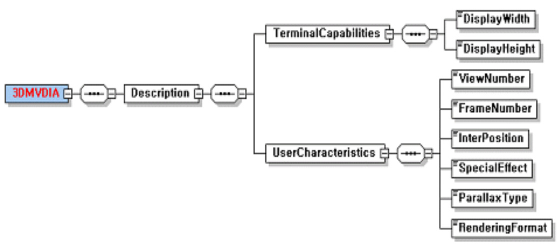

Fig. 2: Schematic diagram of proposed multiview video descriptors under 3DMVDIA DS

UserCharacteristics defines multiview video descriptors used for interactive communication between 3D DIA server and clients. The following descriptors are utilized.

A) *ViewNumber*: selects which view is transmitted. If view *i* is chosen, view *i* and *i*+1 frames, F_i and F_{i+1} are combined to form stereoscopic video and transmitted to a client
B) *FrameNumber*: indicates the number of a current frame.
C) *InterPosition*: indicates the location of a virtual view and is ranged at [0, 1].
D) *SpecialEffect*: includes view switching, sweeping, frozen-moment and rotate.
1) View switching: Users are able to switch flexibly from one view to another as the video continues along time.
2) Frozen-moment and rotate: In the frozen-moment and rotate, time is frozen and the camera view rotates about a given point. One example is that users can view frames $F_1(j)$, $F_2(j)$, …, $F_n(j)$ back and forth at the jth frame of time instant.
3) View sweeping: It involves sweeping through adjacent view direction while the time is still moving. It allows the user to view the event from different view direction. One example is that a user can view frames $F_1(j)$, $F_2(j+1)$, …, $F_n(j+n-1)$ starting at the *j*th frame of the 1st view.
E) *RenderFormat*: Display device can have diverse types of rendering formats such as interlaced, side by side, anaglyph and so forth.

If a user request adapted descriptors to a server, then the server sends adapted video and descriptors.

3. DETAILED IMPLEMENNTATION

3.1. 3D Multiview DIA

The adaptation begins with 4:2:0 YUV format and changes the frame size according to *DisplayWidth* and *DisplayHeight*. IVR generates novel views according to *ViewNumber* and *InterPosition* descriptors. The fast IVR algorithm is presented as follows:

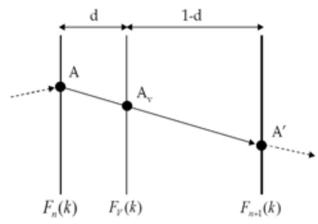

Fig. 3: Two views $F_n(k)$ and $F_{n+1}(k)$ and a virtual view $F_v(k)$. d is the value of *InterPosition.*

A pixel A of $F_n(k)$ is moved to A_v of $F_v(k)$. The translational vector AA' is usually computed by camera transformation matrix. Therefore, during the period that the client requests a service, this transformation can be done and saved. Then $F_v(k)$ can be generated quickly. If the value of *InterPosition*, d between $F_n(k)$ and $F_v(k)$ is determined, the coordinates of A_v can be estimated as follows.

$$A_v = d \cdot \overrightarrow{AA'} \qquad (1)$$

Moving all the pixels of $F_n(k)$ using Eq. (2), $F_v(k)$ is reconstructed.

An algorithm to generate a texture image of a virtual view is carried out as follows:

1) Specify two camera views; cameras 1 and 2.
2) Project pixels from the camera 1 back to 3-D space. For every pixel, its associated 3-D point is estimated by Eq. (1).
3) Project all 3-D points reconstructed in 2) in virtual camera direction and obtain a new image.
4) Repeat 2) and 3) for the camera 2 to fill any holes and update the new image.

3.2. Mobile Client

Considering the low resource and processing capabilities of mobile devices, the client structure needs to be simple. PDA is a device that is considered in our service framework. The PDA is equipped with an user interface software, an H.264/AVC decoder, and a descriptor generator. The descriptor generator produces descriptors in XML containing user preference and terminal capability information.

The following is an example of DIA description of a client in XML.

```
<?xml version="1.0" encoding="UTF-8"?>
<3DMVDIA>
<Description>
<TerminalCapabilities>
<DisplayWidth> 320 </DisplayWidth>
<DisplayHeight> 240 </DisplayHeight>
```

```
</TerminalCapabilities>
<UserCharacteristics>
<ViewNumber> 0 </ViewNumber>
<FrameNumber> 1 </FrameNumber>
<InterPosition> 0.3 </InterPosition>
<SpecialEffect> none </SpecialEffect>
<ParallaxType> negative </ParallaxType>
<RenderingFormat>interlaced
</RenderingFormat>
</UserCharacteristics>
</Description>
</DIA>
```

Considering the low resource and processing capabilities of mobile devices, the client structure needs to be simple. PDA is a device that is considered in our service frame

3.3. Feasibility Validation

We have examined the feasibility of the proposed framework by analyzing the frame rate that a client can process and decode.

Given

- *p*: adaptation and H.264/AVC encoding time in sec
- *c*: the decoding time of a client in sec
- *s*: the frame size in byte
- *n*: frame rate of a server
- *b*: channel bandwidth in bits per sec
- *t*: round trip time (RTT) in sec,

the frame rate *f* is computed by

$$f = \frac{n-1}{p + \frac{ns}{b} + t + c} \qquad (2)$$

For a frame rate *f*, the minimum interval of a server is 1/*f*. the frame interval of a client is then (*n*-1)/*f*. The total response time from the client request to receiving is the sum of server reply time, data transmission time, RTT, and client decoding time. Because the interval between two subsequent frames is greater or equal to the total system response time, Eq. (3) is applied.

$$\frac{n-1}{f} \geq p + \frac{ns}{b} + t + c \qquad (3)$$

4. EXPERIMENTAL RESULTS

For validation of our method, we have used a *breakdancing* multi-view sequence provided by Microsoft Research (MSR). MSR provides camera calibration parameters of eight base cameras and 100 frames of video along with depth data [7].

The time complexity is analyzed. The fast IVR method requires 0.217 sec compared with a general method [8]. The latter needs 1.17 sec. Our method is five times faster. The time was estimated by averaging the processing time of 100 frames of resolution 1024x768. If the image resolution is reduced, much time can be saved.

Fig. 4 shows the variation of a frame rate according to a processing time of a server and a client. If (*p*+*c*) is approximately 3 sec, a client can decode 5 frames per sec. As well, 9 frames can be decoded if (*p*+*c*) is equal to one sec.

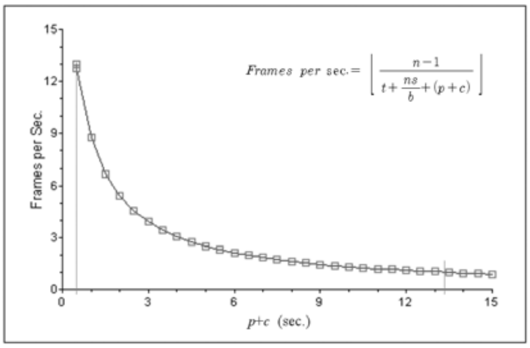

Fig. 4: The variation of the frame rate

Fig. 5 shows a PDA screen displaying stereoscopic video. H.64/AVC encoded data is decoded and rendered. The average frame rate is approximately 10 fps. The stereoscopic video can be observed with anaglyph glasses.

Fig. 5: PDA screen showing stereoscopic video rendered by an H.264/AVC decoder

5. CONCLUSION

In this paper, we proposed multiview video service framework over mobile environments. The main works are fast intermediate view generation and stereoscopic video transmission to compensate for low processing capability of mobile devices. The efficient H.264/AVC codec was used in order to obtain good coding efficiency. The stereoscopic video tested on a PDA produced satisfactory 3D video. Based on proposed MPEG-21 multiview video DS, more efficient interaction between the server and client is possible.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Knowledge Economy, Korea, under the Information Technology Research Center support program supervised by the Institute of Information Technology Advancement. (Grant number GIST-RBRC IITA-2008-C1090-0801-0017)

6. REFERENCES

[1] C. Fehn, R. de la Barré and S. Pastoor, "Interactive 3-DTV-concepts and key technologies", Proceedings of the IEEE, Special Issue on 3-D Technologies for Imaging & Display, 94(3):524-538, March 2006.

[2] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntiens, “A survey of perceptual evaluations and requirements of three-dimensional TV”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 3, pp. 381-391, 2004.

[3] A. Vetro, W. Matusik, H. Pfister, and J. Xin, “Coding approaches for end-to-end 3D TV systems”, Picture Coding Symposium, 2005.

[4] “MPEG-21 DIA AM(v3.0)”, ISO/IEC JTC 1/SC 29/WG 11/N5179, October 2002.

[5] Seungwon Lee, Ilkwon Park, Manbae Kim, Hyeran Byun, “Integrating Multi-view Transmission System into MPEG-21 Stereoscopic and Multi-view DIA(Digital Item Adaptation)”, Proc. of SPIE, Vol 6391 63910Z-1~8, 2006.

[6] Ilkwon Park, Manbae Kim, Seungwon Lee, Young Han Lee, Jun-sup Kwon, Won-Young Hwang, Hong Kook Kim, Jin-Woo Bae, Ji-sang Yoo, and Hyeran Byun, “Interactive Multi-view Video and View-dependent Audio under MPEG-21 DIA”, 3DTV conference 2007, Kos Island, Greece, pp.7-9, May 2007.

[7] http://research.microsoft.com/vision/InteractiveVisualMediaGroup/ 3DVideoDownload/, MSR

[8] L. McMillan, “An image-based approach to three-dimensional computer graphics”, Ph.D thesis, University of North Carolina, 1997.