Multi-view Video Streaming System with MASS library

Shiho Kodera[1]　Takuya Fujihashi[2]　Shunsuke Saruwatari[1]　Munehiro Fukuda[3]　Takashi Watanabe[2]

Shizuoka University[1]　Osaka University[2]　University of Washington[3]

1 Introduction

Multi-view video is one of video technologies to give more immersive and realistic perception to viewers. Figure 1 shows the structure of a future multi-view video streaming system. The multi-view streaming system consists of three sections. The acquisition section captures various scenes by multiple cameras. The transmission section encodes the captured videos and transmits them. In the display section, many viewers watch multi-view video on various devices. Each viewer transmits a request to the transmission section based on own device.

The present paper focuses on the new multi-view transmission section to accept various requests from the display section and video sequences from the acquisition section. When the transmission section receives the various requests, the encoding delay at the transmission section much increases. To decrease the encoding delay, the present paper proposes the parallel encoding system for multi-view video transmission section using MASS library (Multi-Agent Spatial Simulation library)[1]. MASS library is a parallel library for multi-agent spatial simulation for cluster systems.

2 Problem in Multi-view video streaming system

Figure 2 shows the timing diagram of 3-tier multi-view video streaming system. When multiple viewers request multi-view video, the web server encodes the multi-view video for each user. For example, when viewer 1 requests view 1 in video 1, the web server needs to encode the video with high resolution because viewer 1 watches the video on TV.

However, the 3-tire model takes a high encoding cost at the web server. To calculate the encoding cost at the web server, we measured the encoding time. We used JMVC[2] encoder and the video sequence "Ballroom" provided by MERL[3]. We used four types of resolution : 640×480, 352×288, 352×240, 176×144 and 176×120. The encoding time of each video sequence is 320.118 sec (640×480), 112.165 sec (352×288), 92.686 sec (352×240), 23.958 sec (176×144), and 20.705 sec (176×120), respectively. The results show that the encoding time is much long regardless of the resolution. If many viewers request multi-view video with high resolution (640×480), each viewer has to wait 320 seconds.

To encode the multi-view video in real-time, the web server needs the same number of threads for video encoding. In particular, high resolution and low resolution (176×120) video needs 320 and 20 threads, respectively. When the web server uses GPU for the video encoding, one GPU encodes three high resolution video sequences or 51 low resolution video sequences in real-time because one GPU has 1024 threads. References [4, 5] show that the number of Youtube viewers of PCs and smart phones (iOS and Android) in the United States is 128 million/month and 71 million/month, respectively. If each viewer watches the video in 300 seconds, 15000 and 8200 viewers watch video on PC and smart phone per 1 second, respectively. If all viewers request different video, we need 5000 GPUs for PC viewers and 160 GPUs for smart phone viewers. When one PC has two GPUs, about 2600 PCs need for the web server to encode all video sequences in real-time.

3 Multi-view video streaming system with MASS library

To reduce the encoding cost, we propose multi-view video streaming system with MASS library. MASS library is a computer language library to develop and execute multi-agent simulations with multiple computers. In this paper, we use "place" in MASS library. MASS library assigns one process to one "place" and carries out each process at the same time. Our multi-view video system achieves real-time multi-view video streaming by assigning web server processes to each place for parallelization. Figure 3 shows the timing diagram of our system.

1. Each viewer sends the request.
2. MASS library in the web server assures "place 1" and sends the request to the DB.
3. The DB transmits the 1-second video to the "place 1".
4. "Place 1" encodes 1-second video. At the same time, MASS library assures "place 2" and the DB transmits the video of the next 1-second video to the "place 2".
5. "Place 2" encodes the video in parallel.
6. After encoding, "place 1" transmits the encoded video to the viewer. At the same time, MASS library assures "place 3" and the DB transmits the next 1-second video to the "place 3".

Our multi-view video streaming system repeats storing, encoding, and transmitting in each place.

4 Conclusion

In the future multi-view video streaming, many video sequences are captured and many viewers request the video sequences by various devices. To accept various requests, the multi-view video system needs to extend the transmission section. In this paper, we revealed the problem of the current transmission section and proposed multi-view video streaming system with MASS library for the new transmission section.

References

[1] T. Chuang and M. Fukuda, "A parallel multi-agent spatial simulation environment for cluster systems," IEEE CSE, pp.143–150, 2013.
[2] Joint Video Team Of ITU-T VCEG And ISO/IEC MPEG, JMVC (Joint Multiview Video Coding) Software, 2008.
[3] ISO/IEC JTC1/SC29/WG11, Multiview Video Test Sequences from MERL, 2005.
[4] Nielsen TOPS OF 2013, "Digital top 10 u.s. online video brands of 2013," 2013.
[5] Nielsen TOPS OF 2013, "Digital top smartphone apps of 2013," 2013.

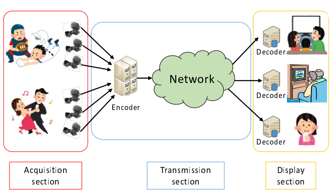

Fig. 1 Multi-view video streaming system

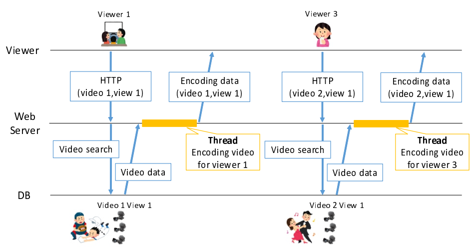

Fig. 2 3-tier model

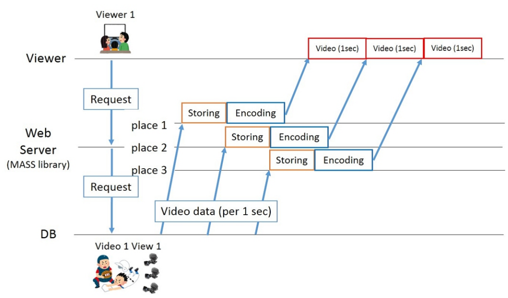

Fig. 3 Multi-view video streaming system with MASS library