Interactive Distribution Technologies for 4K Live Video

Yasuaki Tanaka and Daisuke Ochi

Abstract

Video technology is becoming more sophisticated, and the amount of available content is increasing substantially. Accordingly, individual preferences for video content are becoming more diverse. We have perceived a need for a means of personalized viewing that enables viewers to select their preferred subjects and scenes within the content rather than just passively selecting video content created by third parties, as in the conventional style. This article introduces a system that partitions 4K video into tiles, compresses it, and enables only parts selected by the viewer within the video to be distributed live at the desired size and with high quality.

Keywords: interactive viewing, MVC, video distribution

1. Introduction

The selection of video content available to individual viewers has been increasing recently because of enhancements in video distribution services and the popularity of video sharing sites. Additionally, with set-top boxes and terminals such as smartphones and tablets, viewers are getting used to watching whatever program or video they like at any time. Moreover, the video content produced most recently includes camera work and scene switching that are intended to make the content accessible to the maximum number of viewers. Thus, there are cases in which not all subjects or scenes in a video are shown, even though some viewers might be interested in them. Consequently, it is not necessarily possible to satisfy all viewers.

NTT Media Intelligence Laboratories has been working to resolve this sort of issue and has been conducting research and development (R&D) on interactive viewing technology that enables individual viewers to dynamically select their preferred subject or scene within the content. Thus, viewers can personalize their view of the content to their preferences. This technology enables viewers to change the viewing size or area within the image using pinch or flick operations on a smartphone or tablet (**Fig. 1**).

High-definition (HD) video sources such as 4K and 8K, which are higher definition than conventional HD, are expected to spread quickly in the future. However, on devices with a small viewing area, especially mobile terminals, it will not be possible to provide adequate resolution for viewing, so a viewing style that allows viewers to select the area and field of view themselves will be useful. However, video with 4K class image quality or greater requires an extremely large amount of data, so it is important to find a distribution method that will not overburden the communications bandwidth or terminal resources.

To this end, we have created a compression and distribution technology that can efficiently distribute only selected areas of a video, in order to achieve interactive viewing that enables users to view the specific areas they want to see, for video sources with resolutions of 4K and higher. We have also built an interactive distribution system for 4K live video that can perform all processing on 4K video in real time, from capture through compression and distribution.

2. Technologies for interactive viewing

2.1 Video compression and distribution technology

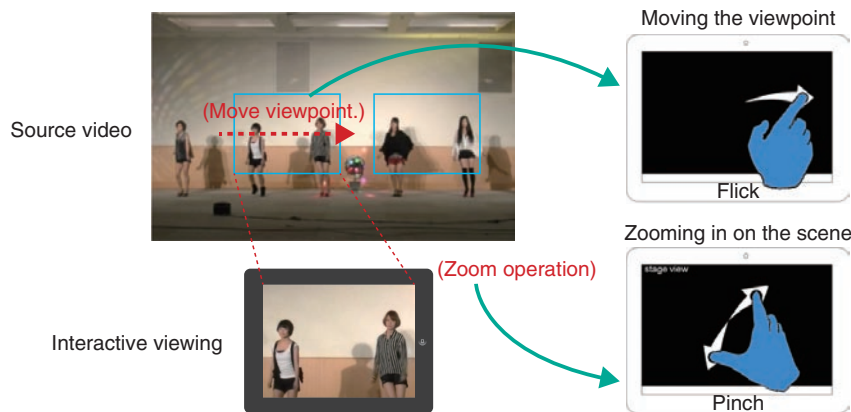The system described here uses H.264/AVC
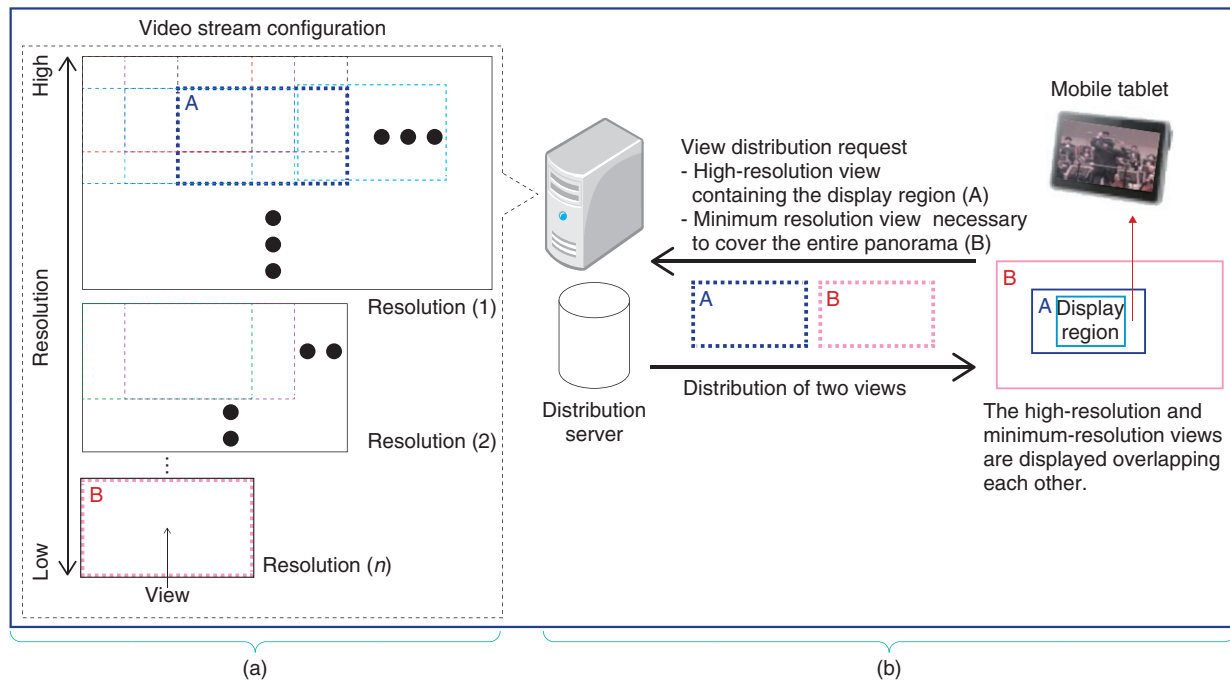
Fig. 1.   4K interactive viewing.

Fig. 2.   4K interactive distribution technology.

(Advanced Video Coding) Annex H multiview video coding (MVC) in order to efficiently distribute the areas selected by viewers. Originally, this method was standardized to handle multiple video streams (e.g., the right- and left-eye images for three-dimensional (3D) video), compressing them together while maintaining synchronization between them. In this case, the standard is used to compress the multiple videos together after partitioning the video source into small tile regions (**Fig. 2(a)**). Specifically, the video source is down-sampled to generate multiple videos at lower resolutions. These multiple videos together with the source video are partitioned into multiple tiles in the compression system. Next, all tile images are encoded in a synchronized compression that is coded into a video stream using MVC. Then, the tiles at each resolution are converted to multi-bit-rate data by coding at multiple compression rates. The
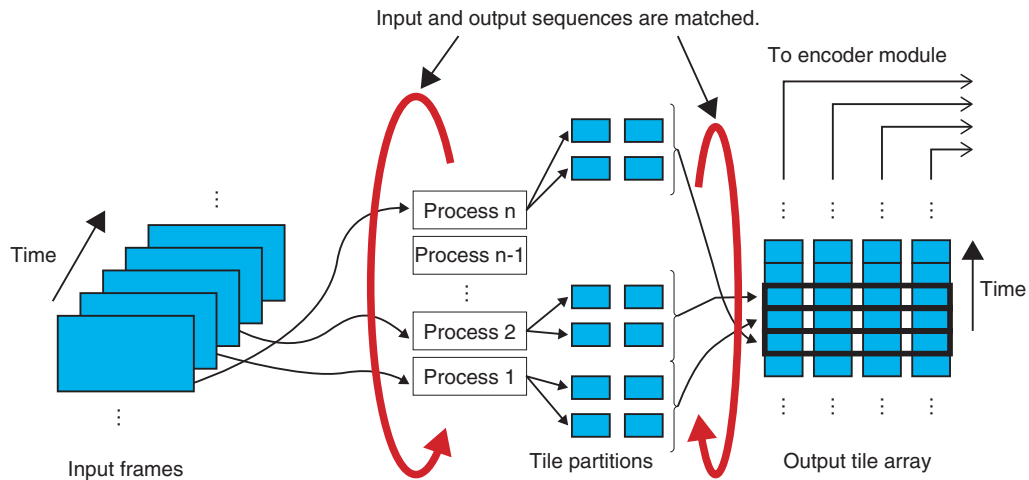
Fig. 3.   Multi-thread processing of tile partitions.

compression-coded video stream is stored on the distribution server, and tile images with the appropriate position and resolution are distributed according to requests from individual viewers.

A distribution scheme is used in which the tiles containing the viewing area selected by the viewer are distributed with high resolution and at a high code rate, and tiles covering the range outside of the screen are distributed at low resolution and at a lower code rate (**Fig. 2(b)**). This is advantageous in that even if the viewer changes the viewing area randomly, the low-resolution tile can be viewed until the high-resolution tile for that area has been requested and delivered. Thus, no areas are dropped from the image, viewers can change the viewing area seamlessly, and network bandwidth is used more efficiently than simply distributing the entire high-resolution video stream.

2.2   High-speed compression technologies for live distribution

To provide live interactive viewing, all tiles generated from the video source must be compressed in real time, which requires more resources and processing time than for generating ordinary streams such as H.264/MPEG-4 AVC. Consequently, the speed of the compression process needed to be increased. The following three techniques were used to increase speed in the current system.
(1)   Pipelining of module input/output (I/O) processing

The compression system in the current proposal requires several modules in sequence for resolution conversion, tile partitioning, encoding, and multiplexing. If the processes in each module are executed sequentially, the total processing time is the total of the processing times for each module, which results in large delays. To reduce the delay, data processing and data I/O for each module were done asynchronously, reducing the cumulative delay. Processing is pipelined within each module into three stages: input, processing, and output, and each stage is processed on a different thread. Data are transferred between threads using data buffers (FIFO[*1] queues) within the application.
(2)   Multi-threaded tile generation

Resolution conversion and tile partitioning of the input video are done in units of video frames. There are no dependencies between these two processes, so an increase in speed was achieved by processing data frames in parallel. This can result in changes in frame order unless the I/O sequence is controlled. Therefore, we used a configuration in which the parallelization processing stage ran on a separate thread, and each thread had a separate FIFO queue. At the input stage, data are placed in each FIFO queue in a simple round-robin[*2] sequence, and at the output queues, data are removed in the same sequence as they were input on the input side (**Fig. 3**). Inputting and outputting data in the same sequence prevents any changes in the order.

---

*1   FIFO (First-In, First-Out): A method of storing data in a location and extracting it in the same order as it was stored.
*2   Round robin: A method of allotting a finite resource by taking turns in order.
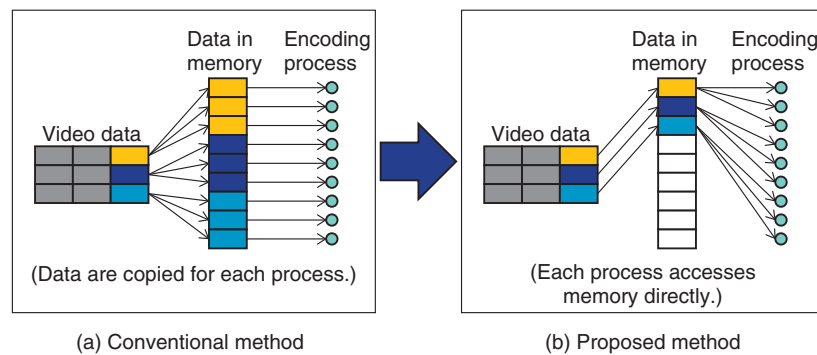
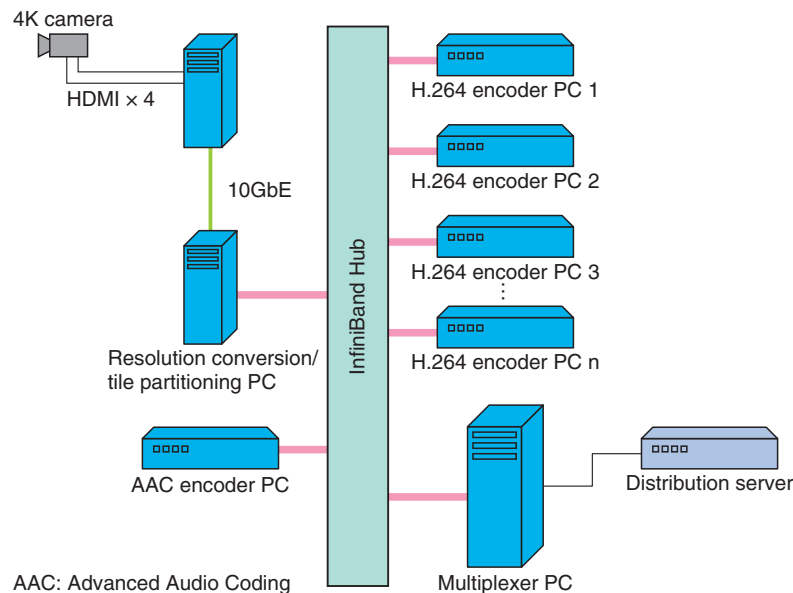Fig. 4. Shared memory for the encoding process.

Fig. 5. System device configuration for live distribution tests.

(3) Handling multiple bit rates with shared memory

When this system encodes a tile at a given resolution and partition position at multiple bit rates, the data are copied to shared memory, and each encoding process accesses the same memory. This allows the tile data to be copied to shared memory only once, reducing the amount of copying done for inter-process communication (**Fig. 4**).

3. Interactive distribution system for 4K live video

The above technology was implemented in a system for interactive distribution of 4K live video, and live experiments were conducted to distribute video from the NTT Musashino R&D Center research facility, where the video was first captured, to the NTT Yokosuka R&D Center, where it was compressed and distributed, and back to the NTT Musashino R&D Center for playback.

An overview of the equipment configuration is shown in **Fig. 5**. Data transmission between the personal computers (PCs) was carried out using an InfiniBand Fourteen Data Rate (FDR) (4X) with theoretical throughput of 56 Gbit/s. There were 22 encoder PCs, and a commercial 4K camera was used for capturing and producing a video source with 4 HDMI (High-Definition Multimedia Interface) lines
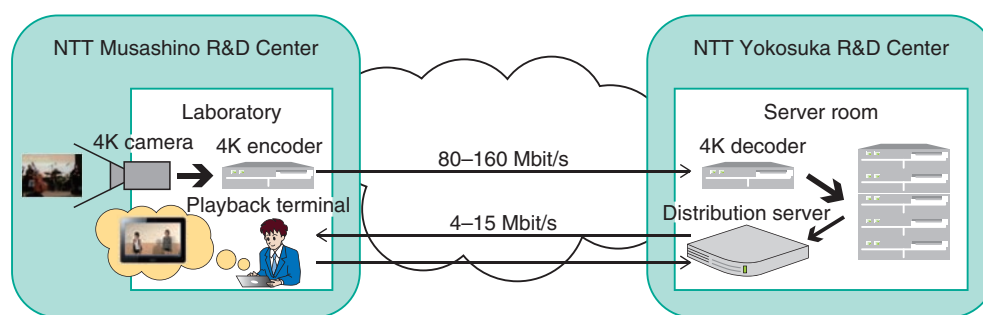
Fig. 6.   Live distribution test configuration.

at 30 fps. This was compressed with MPEG-2 encoders generating total data at a rate ranging from 80 to 160 Mbit/s, which was transmitted to the remote location via a network using a network adapter. After this video was decoded, it was input to the encoder PC for interactive viewing. The video was compression-coded on the encoder PC, then sent in real time to the distribution server. We confirmed that viewers were able to view the video interactively over the network by connecting to the distribution server at approximately 4 to 15 Mbit/s, with a delay of approximately 7 s (**Fig. 6**).

4.   Future prospects

These experiments achieved live interactive viewing, enabling users to view the areas of their choice at high quality from a 4K video source. This distribution and viewing technology is not limited to 4K video; it could also be applied to 8K video, wide panoramas, or whole-sky imagery. We are also conducting R&D in an effort to develop free-viewpoint video distribution, which allows a greater degree of freedom when changing viewpoints. This includes activities toward international standardization of 3DV/FTV (3D video/ free-viewpoint television)[*3].

[*3] 3DV/FTV: International standards for 3D video coding methods by making use of multi-viewpoint video and depth maps under study at JCT-3V, a Joint Collaborative Team on 3D video that includes the international standardization organizations MPEG (Moving Picture Experts Group) and ITU-T (International Telecommunication Union, Telecommunication Sector). Multiple standards are also under study depending on the base compression method used, including 3D-AVC and 3D-HEVC (High Efficiency Video Coding).

Yasuaki Tanaka
  Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.
  He received the B.E. and M.E. degrees in electrical engineering from Osaka University in 1986 and 1988, respectively. He joined NTT in 1988 and studied optical disk systems. He is currently studying live video distribution techniques and 3D video communication systems.

Daisuke Ochi
  Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.
  He received the B.E. degree from the Department of Electrical Engineering and Computer Science and the M.E. degree in information engineering from Kyushu University, Fukuoka, in 2001 and 2003, respectively. His research interests include computer-human interaction, human interfaces, and multimedia information systems.