Technology Reports

VR | Live Video Streaming | FPGA

8K VR Video Live Streaming and Viewing System for the 5G Era

Communication Device Development Department Naoto Matoba

NTT DOCOMO has developed the world's first VR 360° 8K video live streaming and viewing system. This system enables real-time operation through stitching equipment to synthesize video from a number of cameras into 360° 8K video without visible seams and encoding equipment that compresses and uploads the 360° 8K video to a streaming server. Both pieces of equipment use FPGA technology. Also, HMD using Panorama Cho Engine technology and high-resolution liquid crystal displays enable 360° 8K video viewing.

1. Introduction

In recent years, advances in Virtual Reality (VR)[*1] technology have led to the creation of viewing environments with 360° video[*2] that offer high level sensations of presence and immersion. The resolution of cameras capable of capturing 360° video has advanced from full High Definition (HD)[*3] to 4K[*4], while professional cameras capable of capturing 360° 8K[*5] video have also started appearing.

The resolution of Head Mounted Displays (HMDs)[*6] for 360° video viewing also continues to advance, and the fast communications speeds of the coming 5G technologies will enable streaming[*7] and live video streaming of the large amounts of data required for 4K and 8K. Hence, the combination of VR and 5G technology holds promise for services offering high-presence entertainment experiences such as sports and live performances to users in remote locations, as if the viewer is actually in the venue.

Live streaming of 360° video has almost never been attempted, because compared to normal video, the wide area of the 360° video display can suffer from unsatisfactory resolution, which made it impossible to deliver video with a quality good enough

*1 VR: Technology that gives the user the illusion of being in a virtual world. In recent years, this illusion is mainly achieved using HMD (see*6) technologies that affect the user's visual perception.

*2 360° video: Video that covers the entire available field of vision in all directions - front and back, right and left, and up and down.

*3 Full HD: A video screen format consisting of approximately 1,000×2,000 pixels in the vertical/horizontal directions.

to satisfy users.

Also, because it's not possible to capture 360° video with one camera, video signals from multiple cameras must be stitched together (stitching*8), encoded*9 and uploaded in real time. However, real-time stitching and encoding entails very high processing loads, which means so far 360° video has been limited to 4K, even with professional equipment. In addition, delivering 360° 8K video from a server also would require very expensive specialized equipment to decode*10 the 8K video so that it can be viewed, which makes it unachievable within the bounds of equipment configurations available to ordinary users. Additionally, user HMDs do not have sufficient resolution to reproduce 360° 8K video.

Hence, aiming for live streaming of 360° 8K video for the first time, we prototyped equipment to enable 360° 8K video capture, streaming and viewing in real time.

Specifically, we developed (1) stitching equipment to synthesize video signals from multiple 4K cameras and generate 360° 8K video in real time, (2) encoding equipment to compress 360° 8K video and upload it to a server in real time, (3) Panorama Cho Engine®*11 (PCE) technology-based PCE encoding equipment for real-time streaming of 360° 8K video to users in a viewable format, (4) a PCE player to play back the PCE encoded data, and (5) an HMD using one 2K-resolution liquid crystal display per eye. Furthermore, to verify the usability of the equipment, we installed outdoor cameras connected to 5G equipment to test 360° 8K video live streaming.

This article describes the structure of the 360° 8K video live streaming and viewing system we developed, and describes a live video streaming demonstration experiment and its results.

2. Structure of the 360° 8K Video Live Streaming and Viewing System

2.1 System Structure and Objectives

Generally, equirectangular format*12, an intermediate format, is used for 360° video streaming, because of the necessity to convert 360° video to regular video formats to compress and deliver 360° video with existing software and hardware, and because of the simpler player-side processing for viewing with an HMD. Generally, 360° video resolution is indicated by the resolution of the conversion to equirectangular format.

We are aiming to improve the resolution of 360° video using this streaming system to commercialize 360° live video streaming around 2020. Currently, HMD resolution for 360° video is mostly around 1K to 1.5 K per eye. Thus, we assume 2K display resolution per eye will be widely used around 2020.

Comparing the resolution of the HMD display with video resolution and assuming approximately 90° viewing angle in the horizontal direction, 8K resolution with the equirectangular format is necessary to display 360° video on displays with approximately 2K resolution per eye. Therefore, with this development, we set our technical objective as 360° 8K video live streaming.

With 360° video live streaming, video converted to the equirectangular format is transmitted in the same way as normal live streaming, and then played back at the viewer side after conversion for the

---

*4 4K: A video screen format consisting of approximately 2,000 × 4,000 pixels in the vertical/horizontal directions.

*5 8K: A video screen format consisting of approximately 4,000 × 8,000 pixels in the vertical/horizontal directions.

*6 HMD: A device worn on the head that covers the field of vision, and displays a virtual world to the wearer. The wide viewing angle of these devices gives the wearer more intense sensations of emotion and presence.

*7 Streaming: A communication method for sending and receiving audio and video data over the network, whereby data is received and played back simultaneously.

*8 Stitching: Creating 360° video from videos captured with a number of cameras so that the seams between the videos are invisible. The 360° video output after stitching is often displayed in equirectangular format (see*12).

HMD. For this reason, more video processes are involved than normal live streaming and viewing, and these processes have to be done in real time, even though the greater the video resolution, the larger the processing load. Here, "real time" means the time equal to or shorter than the time calculated as the reciprocal of the frame rate (the number of frames displayed per second) for processing one frame*13 of video with the viewing equipment. In other words, for a frame rate of 30 frames per second (fps), processing for each frame of video must finish in each piece of equipment within 1/30 of a second.

Figure 1 shows a schematic of the prototype system we developed. First, video captured with several cameras is combined and converted to equirectangular format, encoding is performed to compress the data into a size that can be transmitted, which is then uploaded to the server. Then, the encoded data is delivered from the server, decoded at the viewer side, and converted to panorama format to be played back on the HMD.

2.2 360° 8K Video Capture, Stitching and Encoding

First, we describe the method of capturing 360°

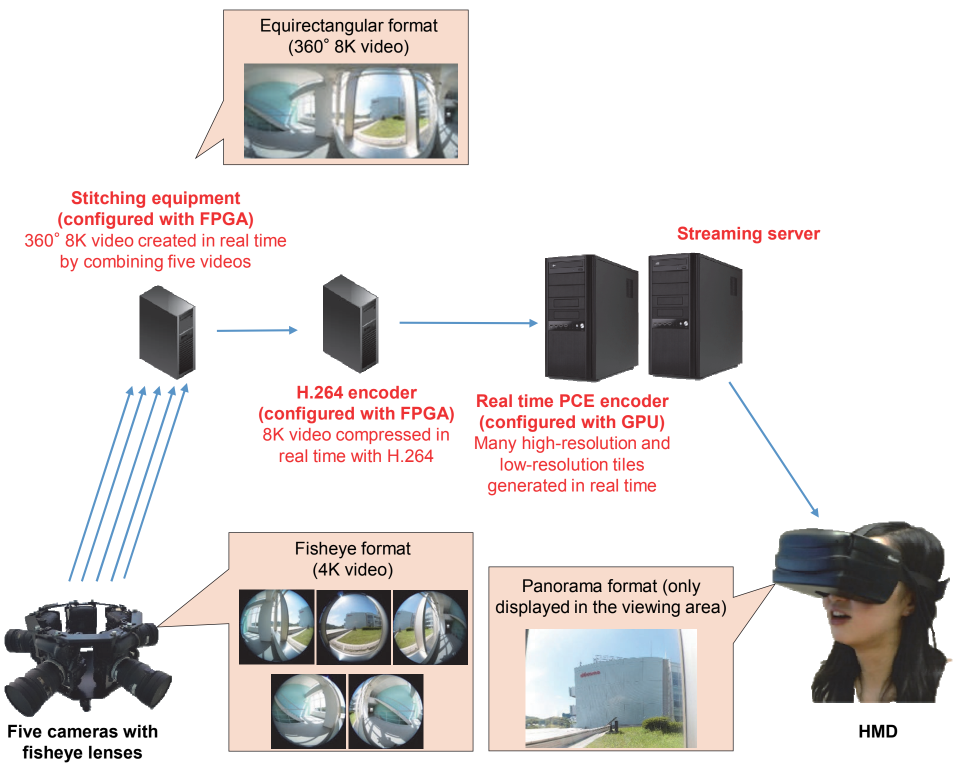

Figure 1 Overall equipment configuration diagram

---

*9 Encoding: In this article, encoding means compressing large amounts of video data so that it can be sent over the network.

*10 Decoding: Reconverting data compressed with encoding equipment back to video data.

*11 Panorama Cho Engine®: Technology that uses streaming technology developed by NTT Media Intelligence Laboratories. A registered trademark of NTT TechnoCross Corporation.

*12 Equirectangular format: A format for projecting spherical 360° video on a flat surface. Normally, this format has a vertical/horizontal ratio of 1:2, and is characterized by projection with latitude and longitude intersecting vertically.

*13 Frame: One of the many single still images that make up video.

video. Because it isn't possible to cover the 360° using a single camera, even with a fisheye lens enabling capture 180 to 200° or greater, multiple cameras are used to capture video covering the 360° [1]. Also, it's not possible to use bulky professional cameras because the distance between cameras is large and stitching becomes impossible. For this reason, a combination of the now common compact 4K cameras is used.

Because the resolution of normal capture in the vertical direction is not sufficient to capture 360° 8K video, we arranged multiple 4K cameras tilted vertically 90° around the circumference and positioned horizontally outward. For the lens, we used a circular fisheye lens capable of capturing a 180° viewing angle. Many of the targets for capture exists close to the horizontal, and since it's not really necessary to install cameras to capture the ceiling or the floor, we only installed horizontally oriented cameras but were able to capture everything up to the ceiling with the fisheye lens. Also, video taken from the center of the fisheye lens with as many cameras as possible would be ideal because of the large distortion around the circumference of the fisheye lens. However, since more cameras means more processing load on stitching equipment due to the increased number of video signal inputs, we used only five cameras in this development.

Next, we described the stitching equipment. Using commercially available equipment for real-time stitching is limited to 4K video output. Currently, no equipment exists that can convert and output video signals with the resolution higher than 4K in real time. Although methods have been suggested that entail multiple stitching of 4K video [2], these methods don't take into account stitching of video from fisheye lenses and would be difficult to apply to this development. **Figure 2** describes the algorithm we used for stitching [3]. First, captured video is rotated 90° because cameras are oriented in the vertical direction, then signals are converted from fisheye format to equirectangular format. Next, by moving the video in vertical, horizontal and rotational directions of the optical axis of the lenses, the five video signals are synthesized and corrected so that video differences between cameras caused by camera fixing jigs are not noticeable. Also, so that the borders between cameras are not apparent, blending processing is performed on multiple parts of the video between the cameras.

Since the stitching processing must be done in real time, we settled on using Field Programmable Gate Array (FPGA)*14 technology because we learnt that it enables high-speed computation processing and can achieve 30 fps.

There are examples of achieving 8K encoding equipment such as [4]. However, considering costs including those of decoding equipment, we decided to use FPGA with H.264*15 IP core*16 to achieve encoding for this development.

A Serial Digital Interface (SDI)*17 is used for transferring video signals between the camera, stitching FPGA and encoding FPGA. With 360° video, it's not possible to position the stitching and encoding equipment close to the camera because the range of capture is in all directions. Also, flexibility is required for arranging equipment for outdoor experimentation to suit the conditions of the site. By using an SDI, it's possible to interconnect equipment using coaxial cable over distances

*14 **FPGA**: An integrated circuit that is configurable after manufacture.

*15 **H.264**: A video data compression encoding method standard recommended by ITU, and widely used in broadcast and Internet streaming, etc.

*16 **IP core**: Partial circuit information summarized as functional units for developing FPGAs, etc.

*17 **SDI**: A widely-used video signal transmission standard, mainly with professional video equipment, and that enables transmission of non-compressed video signals on coaxial cable.

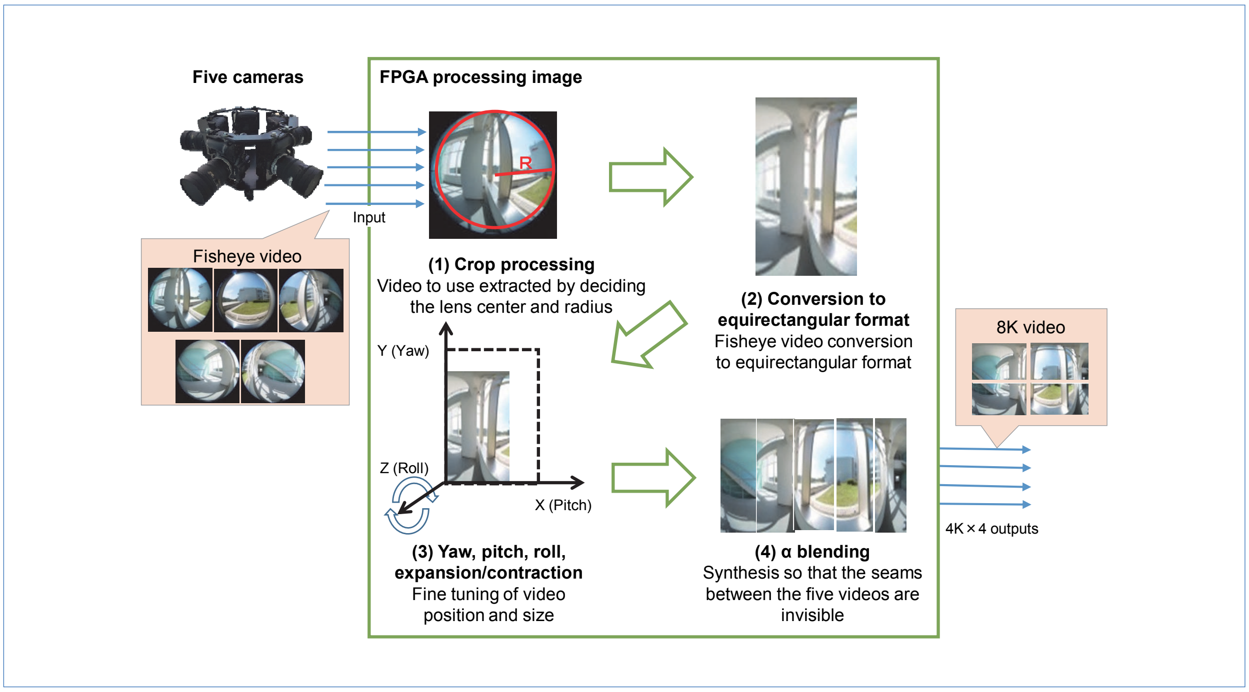

Figure 2 Stitching algorithm

up to 100 m.

Figure 3 shows the external appearance of the stitching and encoding equipment we prototyped.

2.3 360° 8K Video Streaming and Decoding

Next, we describe the method of delivering and decoding the 360° 8K video.

Because of the large load involved with decoding 8K video delivered from the server for viewing at 30 fps, play back in real time is problematic even using a high-end PC with good decoding performance. Also, transmission capabilities of 80 to 100 Mbps are required because the bit rate for encoding must be sufficient to transmit signals without large drops in picture quality. Hence, aiming for commercialization around 2020, both the necessary transmission performance requirements and viewing equipment processing load must be reduced.

We used PCE technology in this development to solve these issues. **Figure 4** describes the principle of PCE. When viewing 360° video with an HMD with the horizontal viewing angle at 90°, the remaining 270° of video cannot be seen. Also, the direction that the user is viewing the 360° video is measured using sensors such as gyroscopes so that only the video in the viewing direction is displayed after decoding. Here, with PCE, the data for the detected viewing direction is sent to the streaming server, and the video in that direction is sent with the same resolution as 360° 8K video (a high-resolution partial tile). In addition, considering that the user might move his or her head suddenly, the 360° video is simultaneously sent with the resolution

Figure 3　Stitching equipment (left), encoding equipment (right)

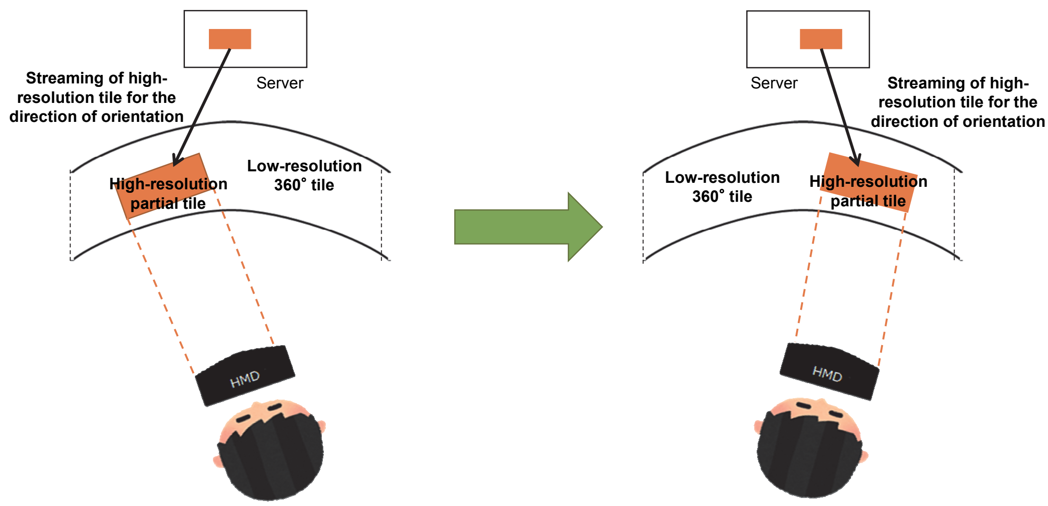

Figure 4　PCE principle

reduced (a low-resolution 360° tile). With this system, the video is always 8K resolution in the line of sight of a static user, but maintained at a lower resolution when the user suddenly moves his or her head.

In this development, considering the viewing angle, we used 2K for the resolution of both the high-resolution partial tiles and the low-resolution 360° tiles. This enables the decoding load to be two videos at 2K resolution, which is a significant reduction in processing load compared to 8K video, and enables play back through an HMD using processing capabilities equivalent to the latest smartphones or similar devices.

As it's necessary to generate these tiles in real time at the server, in this development, 16 high-resolution partial tiles in various directions and one low-resolution 360° tile are generated in real time in consideration of processing load at the server.

PCE technology enables viewing of video with the same quality as 8K.

2.4 Viewing with HMD

HMDs used for viewing 360° video consist of displays for the video, lenses adjusted for wide viewing angle focus and a housing designed to maintain suitable display and lens positioning when worn on the head. As discussed, as there are no products with displays suitable for viewing 360° 8K video available on the market, this is a newly prototyped system.

This HMD (**Figure 5**) uses a 2K-resolution liquid crystal display per eye matched to the resolution of 360° 8K video. The pixel density is 1,008 pixels per inch (ppi)*18.

Both displays and lenses influence the quality of the video experienced when viewing with an HMD, but since there were no suitable lens combinations available, we designed and developed a new lens to suit the liquid crystal display.

To raise the sense of immersion and presence, requirements for the lens for this HMD include widening the viewing angle, making the lens small and light for mounting on the head, and minimizing peripheral aberration of the lens.

In this development, we used plastic to design a highly refractive aspheric lens to reduce weight and rectify aberration in the lens periphery. Although combining more of these lenses further improves peripheral aberration, we used three to strike a balance between the overall weight of the HMD and the extent of aberration improvement.

Figure 5　Prototype HMD

3. Verification Experiments of 360° 8K Video Live Streaming Using 5G Equipment

As stated, transmission speeds required to send 360° 8K video will be possible in the 5G era, meaning there is anticipation for high quality 360° video live streaming services.

To verify this technology, we performed a demonstration experiment of 360° 8K video live streaming using 5G at the Niigata Soh Odori event held on September 16, 2018 as part of the Niigata City demonstration experiment project. **Figure 6** describes

*18　ppi: The number of pixels per inch.

Figure 6 The scene at the Niigata Soh Odori demonstration experiment

this demonstration experiment.

During the experiment, we successfully provided stable live streaming of a 360° 8K video experience through HMDs to approximately 400 visitors over almost a full day.

4. Conclusion

We developed a prototype 360° 8K video live streaming and viewing system as an initiative to improve the quality of 360° video streaming for the 5G era. We also conducted outdoor technical demonstration experiments using this prototype equipment. This enabled us to garner a wide range of knowledge about the achievability and effectiveness of 360° 8K video live streaming and clarify a range of issues towards commercialization. Firstly, in solving these issues, we will continue to work on development for even higher quality to improve the user experience. At the same time, we also plan to study how to make the overall system cheaper.

REFERENCES

[1] M. Shohara, H. Sato and K. Yamamoto: "Special edition A: A New Audio Visual Experience, Chapter 2 360° Capture," Journal of the Institute of Image Information and Television, Vol.69, No.7, pp.652-657, Sep. 2015 (In Japanese).

[2] T. Sato, K. Namba, M. Ono, Y. Kikuchi, T. Yamaguchi and A. Ono: "Surround Video Stitching and Synchronous Transmission Technology for Immersive Live Broadcasting of Entire Sports Venues," NTT Technical Review, Vol.15, No.12, Dec. 2017.

[3] Kolor: "Autopano - Tutorials - Quick Start Guide." http://www.kolor.com/wiki-en/action/view/Autopano_-_Tutorials_-_Quick_Start_Guide

[4] Y. Nakajima, Y. Nishida, M. Ikeda, K. Nakamura, T. Ohnishi, T. Sano, H. Iwasaki and A. Shimizu: "Development of 8K video encoder with HEVC realtime encoder LSIs," Vol.40, No.35, BCT2016-76, pp.13-16, Oct. 2016 (In Japanese).