

Real-time Omnidirectional Video Streaming System

*Daisuke Ochi, NTT Media Intelligence Laboratories
Shinnosuke Iwaki, Dwango Co., Ltd.*

Abstract

NTT Media Intelligence Laboratories and Dwango Co., Ltd. have developed a system that makes it possible to view highly immersive video of live events captured by 360-degree omnidirectional camera using an interactive video distribution technology developed by NTT Media Intelligence Laboratories. Dwango has integrated this system with their *Niconico Live* broadcasting system and is now offering a real-time streaming service. We report the details of this service in this article.

Keywords: omnidirectional video, head-mounted displays, video streaming

1. Introduction

In the field of virtual reality (VR), simulating physical presence and participation at events in remote locations has proved to be a challenging task that VR programmers and engineers have long grappled with. There are many diverse elements necessary for this kind of simulation, including not only the five senses but also ambience and other nonverbal sensations, the perfect reproduction of which is considered to be difficult with current technology.

At NTT Media Intelligence Laboratories, we have decided to focus on issues related to the sense of sight and are tackling the simulation of physical presence through video consumption. Until now, the laboratories' accumulated knowledge and know-how related to interactive video distribution technologies have been directed at distributing video to smartphones and tablets with limited network bandwidth, allowing viewers to immerse themselves in remote events of their choosing [1, 2].

Dwango Co., Ltd., on the other hand, has not only been overlaying comments over videos and in other ways providing brand new viewing experiences with its user-uploaded video sharing service, *Niconico Live*, but it has also made several attempts to share real-world immersion with Internet users through actual events it runs such as *Niconico Chokaigi* as

well as event venues that it operates such as *nicofarre* in order to achieve the company's motto of *blurring the borders between the Internet and the real world*.

We have built a real-time omnidirectional video streaming system by using the interactive video streaming technology developed by NTT Media Intelligence Laboratories and taking advantage of the expertise in integrating Internet and real-life experiences cultivated by Dwango. Users of this system can feel as if they were looking at their surroundings while being at the location of a remote event. In this article, we provide an overview of the interactive video streaming technology, describe the techniques employed to build the system, discuss the content design, and introduce topics related to the actual usage.

2. Approach

We make use of two devices in the pursuit of visual immersion: a 360-degree omnidirectional camera that can record video in all directions, and a head-mounted display (HMD) that can track the orientation of the wearer's head. The former is used to record omnidirectional video that captures the sense of being at a (remote) location, and the latter is used to show highly immersive video that responds to the direction in which the viewer is facing [3, 4].

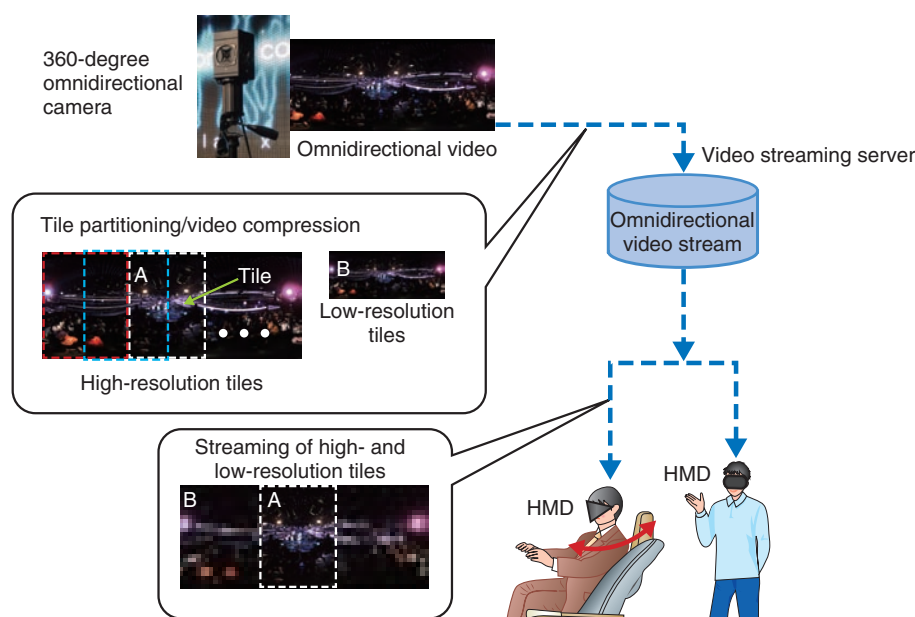


Fig. 1. Overview of interactive video distribution technology.

These sophisticated devices allow viewers to watch highly immersive video, although a broadband network connection must always be available for remote viewers to enjoy clear omnidirectional video that has 4K or higher resolution. However, by applying the interactive video distribution technology [2] developed by NTT Media Intelligence Laboratories and only streaming high-quality images that the viewer is looking at over limited network bandwidth, we have made it possible to create highly immersive experiences even on ordinary networks that are common today, as shown in Fig. 1.

3. Interactive video streaming technology

We use interactive video streaming technology that divides pictures captured over a wide viewing angle into several overlapping tile-shaped regions, compresses each of those tiles at high resolution, and then sends only the high-resolution tile that the viewer is looking at. As shown in Fig. 1, this can reduce network bandwidth requirements because viewers no longer need to receive high-resolution tiles for regions they are not looking at. Furthermore, by compressing the entire video into low-resolution tiles that are sent to the viewer along with the high-resolution tile, we can provide it for the viewer while waiting for high-resolution tiles to load when they look at a different area of the video.

This technology allows viewers to watch immersive, wide-angle video while requiring less network bandwidth than an entire video streamed at high resolution.

4. Implementation of a commercial service

We implemented a commercial system by applying the interactive video streaming technology to omnidirectional video, as shown in Fig. 2. Our goal was to provide a system with an enjoyable, high-quality viewing experience. To do this, we did not simply implement the aforementioned technology; we also incorporated techniques for preserving the quality of long viewing sessions and techniques for keeping viewers engaged, as well as the culture and expertise that Dwango has accumulated over the years, into both the server and client sides of the system. In the following sections, we describe more detailed specifications related to the system's implementation and the techniques for incorporating them.

4.1 Server side

Omnidirectional video was captured using the Ladybug3 omnidirectional camera (Point Grey Research, Inc., Canada) at a resolution of 1920×960 -pixels and a frame rate of 29.97 fps. As shown in Fig. 1, the captured video was compressed in real time into seven 480×960 -pixel high-resolution tiles

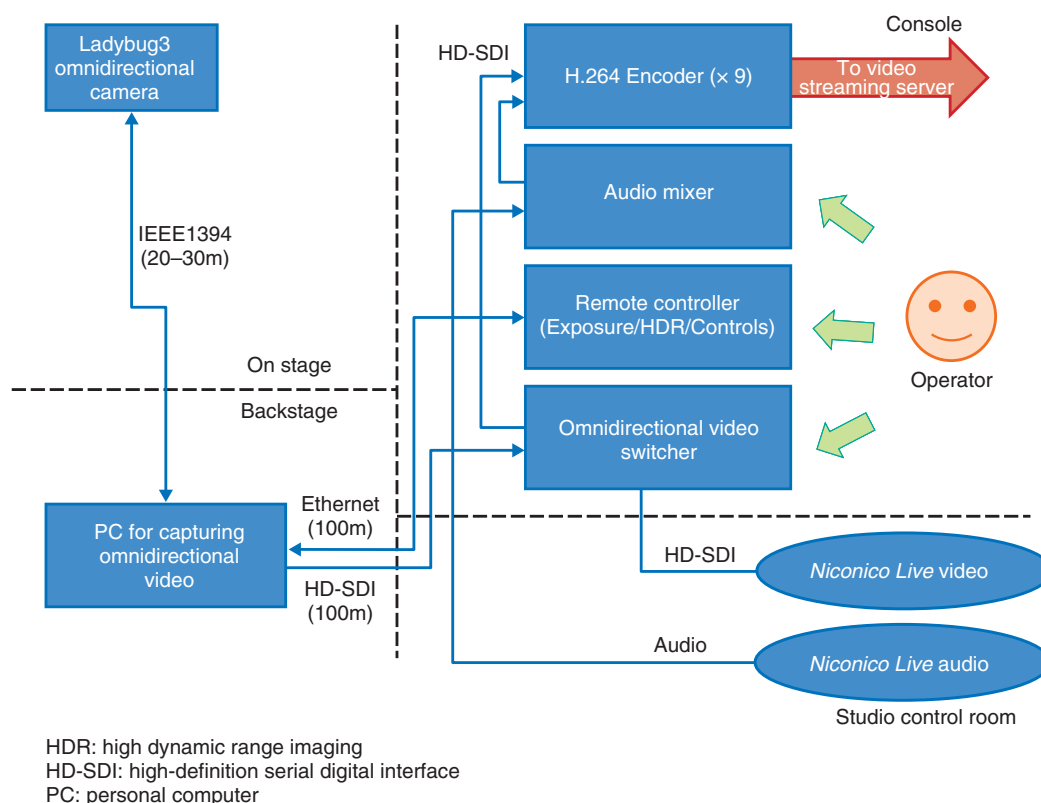


Fig. 2. Overview of commercial system.

(at 1.2 Mbit/s each) and one 512×256 -pixel low-resolution tile (at 0.5 Mbit/s). The encoded tiles are then sent to the distribution server via the Real Time Messaging Protocol (RTMP) and finally streamed to users over a content delivery network (CDN). Video is ordinarily streamed at a sustained bit rate of 1.7 Mbit/s (for one high- and low-resolution tile), but it can temporarily burst up to 2.9 Mbit/s (for two high-resolution tiles and one low-resolution tile) when the viewer turns to look in another direction because it will be sent along with the formerly viewed high-resolution tile. It currently takes about five seconds for encoding, streaming, and buffering with the viewer's application.

We actually ended up using 3.2 Mbit/s of network bandwidth to stream including a video taken by a professional camera operator (described below) at the same time as the omnidirectional video. This reduced bandwidth by approximately 40% over sending the full omnidirectional video at the same quality as the high-resolution tiles (a total of 5.1 Mbit/s, including 0.3 Mbit/s for conventional video and 4.8 Mbit/s for four high-resolution tiles at 1.2 Mbit/s each). There

should be sufficient network bandwidth to handle this in modern homes with fiber-optic connections and on cell phones with LTE (Long Term Evolution) connections. Viewers can also watch highly immersive, time-shifted video on demand because all of the captured omnidirectional video tiles are stored as video files on the video streaming servers.

4.2 Client side

Although there are high hopes for HMDs in the field of virtual reality, they are not yet in widespread use by the general public. Therefore, we developed a client application for web browsers as well as HMDs to allow viewers to enjoy omnidirectional video even without an HMD, as shown in **Fig. 3(a)**. The web-based client allows users to choose a viewpoint by dragging the mouse cursor. As shown in **Fig. 3(b)**, the HMD client allows users to immerse themselves in omnidirectional video once they have plugged their HMD into a personal computer and downloaded the client application. The Oculus Rift (Oculus VR, LLC, USA) is currently the supported HMD; with the gyro sensors installed in the Oculus Rift, the client

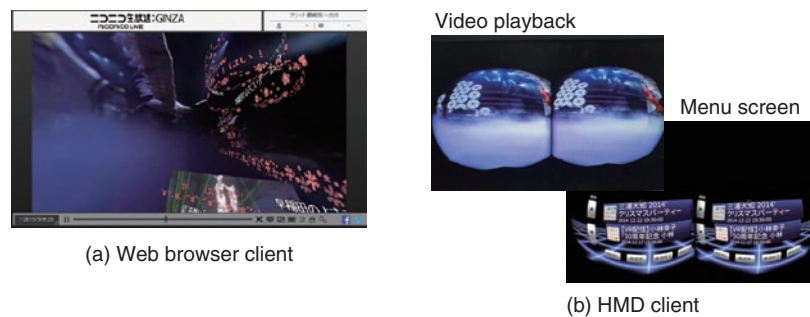


Fig. 3. Client applications.

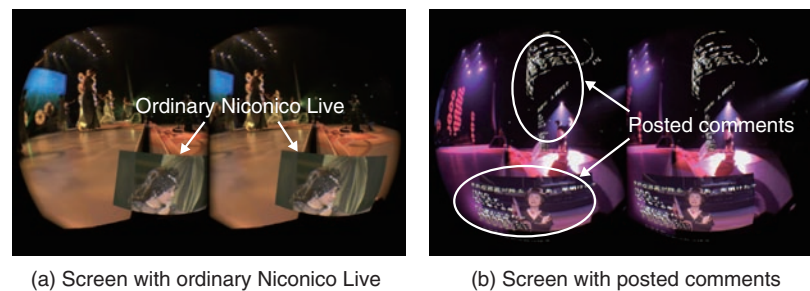


Fig. 4. Monitor placement.

application can detect which direction the user is facing and then request the corresponding high-resolution tile from the streaming server. High-resolution and low-resolution tiles are received by the Adobe Air^{*} runtime system via RTMP and then combined and rendered.

Considering the fact that remote events can continue for long periods of time, we have allowed users who are tired of looking around in the VR environment to watch video (ordinary Niconico Live) that is filmed by camera operators and placed in the omnidirectional video such that it appears to be on a flat monitor, as shown in **Fig. 4(a)**.

Similarly, as shown in **Fig. 4(b)**, we have preserved the fun of using Niconico by floating Niconico's characteristic comments, which allow users to exchange their feelings and impressions with each other, in the same three-dimensional (3D) space as the omnidirectional video. These comments are actually shown with a stereoscopic 3D effect so that they induce the feeling of 3D even though the omnidirectional video was filmed in 2D.

5. Results

This system was used for the first time at the Nippon Budokan on November 17, 2014, during a 50th anniversary event featuring the singer Sachiko Kobayashi (**Fig. 5**). There were no problems during the 185-minute performance. Omnidirectional video of the event was viewed 10,022 times by 4762 separate users via the HMD and browser clients, reaching approximately 10% as many users as those who viewed the ordinary Niconico Live alone. These results allowed us to confirm our expectations with respect to the users who would watch highly immersive (omnidirectional) video. We also learned that it would take longer than the aforementioned five seconds to stream high-resolution tiles for time-shifted video. During long-running performances, the video files accumulated for each tile on the distribution server can grow quite large and thus slow down seeks to arbitrary times in the video stream. We are working on improving the system to solve this problem.

* Adobe Air is a registered trademark of Adobe Systems Incorporated.



Fig. 5. Sachiko Kobayashi.

6. Future work

In addition to solving the problems discussed earlier in this article, we also plan to improve the quality of the video that is distributed and further reduce bit rates. Looking ahead to 2020, we plan to expand into

sports content and explore the use of sound in pursuit of even more immersive experiences.

References

- [1] H. Kimata, D. Ochi, A. Kameda, H. Noto, K. Fukazawa, and A. Kojima, "Mobile and Multi-device Interactive Panorama Video Distribution System," Proc. of the 2012 IEEE 1st Global Conference on Consumer Electronics, pp. 574–578, Tokyo, Japan, Oct. 2012.
- [2] Y. Tanaka and D. Ochi, "Interactive Distribution Technologies for Live 4K Video," NTT Technical Review, Vol. 12, No. 5, May 2014. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201405fa5.html>
- [3] D. Ochi, "Omnidirectional Video Streaming System with HMD," Proc. of the 21st International Display Workshops, Niigata, Japan, Dec. 2014.
- [4] D. Ochi, S. Iwaki, Y. Kunita, J. Hirose, K. Fujii, and A. Kojima, "HMD Viewing Spherical Video Streaming System," Proc. of ACM Multimedia 2014, pp. 763–764, Orlando, USA, Nov. 2014.

See for yourself

If you have a premium Niconico account, you can watch the omnidirectional video recorded for this article on demand at <http://live.nicovideo.jp/gate/1v199935215>. Viewers with an HMD can download the HMD client application from <https://www.dropbox.com/sh/v3tbi6u4vfze3r2/ACaGb99KeLXvgVsugMXRnaQa?dl=0>. Viewers without an HMD can watch the video from their web browser.



Daisuke Ochi

Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.

He received the B.E. from the Department of Electrical Engineering and Computer Science and the M.E. 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.



Shinnosuke Iwaki

Engineer, Dwango, Co., Ltd.

He created *Keitai Douga Henkan-kun* (<http://mobilehackerz.jp/>) and also developed THETA applications such as "THETA physical remote shutter" and "Oculus THETA," and organized THETA hacking information. When he was broadcasting Niconico Live streaming as a user, he was recruited by Dwango's staff with comments flowing on his Niconico Live screen. As an engineer at Dwango, he is responsible for developing the full range of systems to blur the borders between the Internet and the real world, including the production system of nicofarre, a live house surrounded by 360-degree LED displays.