Future-oriented Sports Viewing Project

KOSHIBA Keiichi KINOSHITA Shingo

Abstract

With the Olympic Games, new visual technologies such as color broadcasting, satellite relay and high-definition television have been tested and put to practical use. This games were held under the pandemic disaster, and the challenge was whether a live broadcast could provide the sense of presence and togetherness at the actual venues.

The "Future-oriented Sports Watching Project", in collaboration with partner companies NTT and Panasonic, used dome video streaming and Kirari! Technology to provide a live broadcast of the tension of the athletes at that moment, the power and beauty of their bodies, the sense of unity between spectators and athletes and the unique Olympic atmosphere at the venues. This is the first-ever challenge to convey a sense of reality as it is in real life was undertaken. (Researched by Tokyo 2020 (cooperation of Prof. Okyudo of Wakayama University, an advisor to the Committee) and NTT).

Keywords : Dome image, High realistic, Kirari!

1. Introduction

This project aims to deliver experiences as if one were in the competition venue to someone who cannot visit the venue. The technologies employed were, first, technology that expresses the sense of presence of the venue, which cannot be expressed on a flat screen, by projecting it onto a dome screen, and second, NTT's ultra-realistic communication technology, which projects the players as they really are. This article summarizes the results of the "Future-oriented Sports Viewing Project" implemented at the Olympic and Paralympic Games Tokyo 2020 (hereinafter referred to as "Tokyo 2020 Games"), with the above objectives in mind.

KOSHIBA Keiichi
G1 company Inc.
E-mail kossi@genki-1.com
KINOSHITA Shingo
NTT Human Informatics Laboratories, NIPPON TELEGRAPH AND TELEPHONE CORPORATION. CORPORATION
E-mail shin.kinoshita@hco.ntt.co.jp
The Journal of The Institute of Electronics, Information and Communication Engineers, Vol.105, No.8, Supplement, pp.203-210, August 2022

2. Immersive LIVE Viewing (Dome Video Streaming)

To achieve the vision of the Tokyo 2020 Games to be "the most innovative games in history", we set a target measure that would remain as a legacy after considering innovations in the field of Audio & Video.

There are approximately 350 planetariums in Japan, and no other country in the world has as many dome screens for the public as Japan. Recently, instead of the conventional proprietary system for astronomical representation, there has been a trend to replace the projection system with a high-luminance projector. This system was recommended by Prof. Okyudo of Wakayama University for sport viewing[1].

As a result of our research for various realistic image expressions, we thought that image expressions that allow spectators to share the excitement and wonder of the game with each other are more promising as a new style of sport infection in the future than methods of individual enjoyment using VR glasses, and the like. After many tests and negotiations with the support of Prof. Okyudo, we decided to plan, promote and operate the "Future-oriented Sports Viewing Project : Immersive LIVE Viewing".

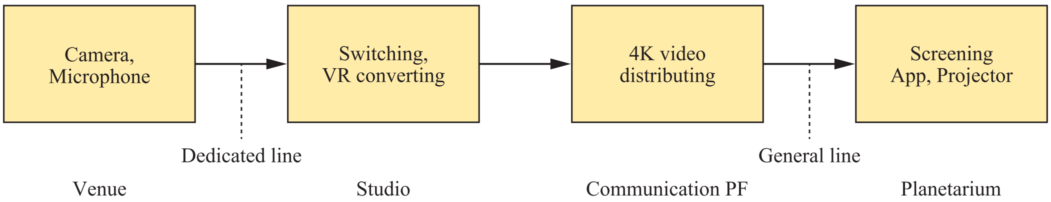


Figure 1　Overall Structure of the Distribution System

The concept of this project was "to have spectators experience a sense of presence and unity each other in a safe", secure and closer location. In this project, the goal was to express the speed, height, and strength of sport by projecting images of competitions on a 10 m class dome screen and to express the sense of presence and unity at the venue by installing cameras in the spectator's seats. We thought that people can enjoy the games with those close to them in the domes and cheer together with others in the venue.

2.1　System Configuration

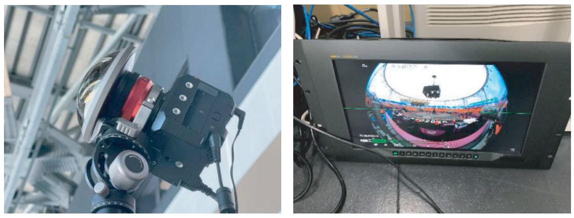
The overall configuration is shown in Figure 1. The system is divided into the following sections depending on the location : venue-dedicated line-studio-distribution platform-general line-screening theatre. In this project, the key point was to ensure that the system would remain as a legacy in the future, i.e., to combine and utilize existing infrastructure and technologies with as little operational cost as possible.

2.1.1　Camera System

Considering that the screen is a hemispheric dome, we decided to use a consumer-use 4K-resolution mirrorless single-lens camera with a fisheye lens. The focal length of the lens was chosen so that the fisheye image is inscribed in the vertical position of the 4K format to achieve the highest resolution possible (Figure 2).

2.1.2　Distribution System

The video format used for transmission was the 16 : 9 4K video format also used in television broadcasting. This format doesn't need to construct a special system. In addition, considering the cost, ease of operation, and stability of the system, we proposed using a best-effort general optical fiber line instead of a dedicated line. Furthermore, we decided to use an existing Internet TV platform to distribute live and recorded images. Thanks to this, the screening operator at each planetarium can select the program menu in STB by remote controller just as if operating a television then operational problems has minimized.



(a)　SLR camera with a fisheye lens　　(b)　Video signal image

Figure 2　Single-lens Reflex Camera with Fisheye Lens and Image

2.1.3　Screening System

Originally, planetariums projected point images of stars on a hemispherical dome screen by means of specialized planetarium equipment. In recent years, however, projectors have made it possible to project images over the entire dome screen so that pictures representing constellations and simulated images for astronomy can be projected. The resolution is improving year by year, and facilities that can project dome images of 8K in diameter using multiple 4K and 8K projectors have been launched. Currently, the systems vary widely depending on the facility and the manufacturer. However, it is possible to project fisheye images (dome master format) with a viewing angle of 180 degrees inscribed in a square, although the maximum possible resolution differs depending on the system.

Therefore, in cooperation with the manufacturers of the planetarium systems installed in each of the planetarium facilities participating, we proposed the addition of a system that can cut out and project fisheye images inscribed at the top and bottom of a 16 : 9 4K image in dome master format.

2.2　Outline of the Experiment

The outline of the experiment is as follows.

2.2.1　Schedule

The screening methods include live transmission and reception with video-on-demand (VOD) of recorded images. As a result, it was decided that the shooting dates were from July 23 to August 8, 2021, and the screening dates were from July 27 to August 8, 2021.

2.2.2 Competitions

The selected competitions were ceremonies, volleyball, 3×3 basketball, badminton and sport climbing. Three cameras (four cameras only for the opening and closing ceremonies) were installed at each competition venue to realize the concept of providing a realistic feeling of being at the spectator's sheets.

Audio equipment was installed in each camera and signals were mixed out.

2.2.3 Camera Locations

It was decided to place the cameras in the spectator's seats rather than on the field to reproduce the concept as much as possible. As a result of negotiations with the international federations and the like, it was decided to film the opening ceremony, volleyball, badminton, 3×3 basketball, and sport climbing from three to four camera locations in the spectator stands. Furthermore, for sport climbing, in which participants climbed a high wall, it was decided to install one camera close to the competition stage since it was confirmed from pre-test filming that the wall height could be expressed with close to the stage.

2.2.4 Program List

We decided a competition schedule to be filmed based on considerations from the spectators' perspective, such as the presence of Japanese national team athletes and medal games, and operational considerations.

2.2.5 Screening Theatres

The final screening schedule was left to each municipality and screening theatre. As a result, a total of 11 live screenings and 44 VOD screenings were conducted. All scheduled programs were filmed, transmitted and screened without any major problems, even though communication methods and procedures for troubleshooting had been prepared in advance. The facilities of the participating planetariums all differed in the resolution of the dome master diameter from 2K to 8K, the dome diameter from 9 m to 25.6 m and the dome screen from horizontal to tilted (Table 1).

2.2.6 Online Survey

An online survey was conducted at the screening theatres under the initiative of Wakayama University to clarify the realism of this experiment. All screening theatres distributed promotional flyers to visitors and asked them to access a web page using a smartphone or other device to answer questions via a two-dimensional barcode printed on the flyer. Iida City Museum of Art conducted its independent open-ended questionnaire survey and exit interviews, and Wakayama University is planning to conduct research and analysis in the future.

2.3 Results

The original plan was to hold public viewing at 11 venues but there was a nationwide movement to refrain from public viewing due to the coronavirus disease 2019 (COVID-19), then resulting in 4 venues that could attract the general public, 4 venues that became limited events for related parties and 3 venues that were canceled. As a result, 969 people (851 for the general public and 118 for related parties) participated in this demonstration experiment.

Although subjective, we have analyzed and organized the three realistic sensations obtained through this

Table 1 List of Theatres

Location		Capacity	Dome	
Planetarium	Prefecture	Seats	Diameter	Type
Katsushika City Museum	TOKYO	172	18 m	Tilted
Fuchu City Museum	TOKYO	218	23 m	Horizontal
Iida City Museum	NAGANO	90	12 m	Horizontal
Fukui City Museum of Astronomy (Seiren Planet)	FUKUI	160	17 m	Horizontal
Tsukuba Expo Center	IBARAKI	232	25.6 m	Tilted
Fukuoka City Science Museum	FUKUOKA	220	25 m	Horizontal
Ishigakijima Hoshinoumi Planetarium	OKINAWA	46	9 m	Tilted
Miraikan (The National Museum of Emerging Science and Innovation)	TOKYO	112	15.24 m	Tilted

experiment as below.

2.3.1 Realism of Sport

We believe that the most significant advantage of images projected on a dome screen is its expressive power that encompasses the entire viewing angle of the spectator. One of features is the expressive power for vertical movement, which is different from the ordinary TV broadcasting.

For example, the athletes who dash up in sport climbing with the fastest speed in the world and all of the trajectory of the badminton shuttle of lob serve above the heads of athletes that takes over two seconds can be followed perfectly. This system's ability to experience sport as they are is a significant feature.

On the other hand, the ordinary 16 : 9 video can be dramatically directed to a high degree of perfection and resolution by zooming and switching. But the dome screen video requires high-performance cameras and lenses and expensive infrastructure to deliver images of the entire surface (hemisphere) in high-resolution. Therefore, the camera location that can deliver powerful images without its zoom function is very important. It is important to consider this point when installing cameras for similar initiatives in the future.

2.3.2 Realism of the Venue

The enjoyment of watching sport is not limited to the sport itself. Especially in the Olympic Games, competition venues are beautifully decorated without advertisements and with many flags of participating countries under the concept of "Clean Venue". This time, by setting up cameras near the spectator seats, we decided to provide "memories of being there" by recording the atmosphere of the entire venue, the works of the event staff and everything else (Figure 3).

Similarly, by recording the sounds from the venue with microphones without adding the running commentary, we could provide a surround sound experience that makes the most of the 5.1 channel, even though it picked up various noises, such as reverberations.

2.3.3 Realistic Sensation of Entertainment

For the opening and closing ceremonies and sport presentations between games, we could convey a unique sense of scale by capturing images of the entire competition venue from a wide-angle, including fireworks, paper airplanes descending, and lighting effects.

However, since the event was held without spectators, it was not possible to verify the effect of the sense of unity between the cheering spectators at the competition venue and the cheering audiences at the screening theatre.

Figure 3 View of the Screening Venue (Location : The National Museum of Emerging Science and Innovation (Miraikan) Photo courtesy of Panasonic) (© 2021-International Olympic Committee-All Rights Reserved)

3. Transmission of Holographic Images of Badminton Matches

We carried out technology demonstrations aimed at delivering the experience of being at the badminton venue of the Tokyo 2020 Games to those who could not attend the actual event by using Kirari! ultra-realistic communication technology to transmit holographic images of the event. This section presents results of demonstrations of a next-generation immersive technology for providing a sense of "being there", namely through the sense of presence and sense of unity, by transmitting holographic images of the badminton matches.

We initially planned demonstrations of live viewing by inviting the general public. However, to prevent the spread of the COVID-19, we cancelled the public demonstration and instead opened "Future-oriented Sports Viewing Project" at the National Museum of Emerging Science and Innovation (Miraikan) (Figure 4).

3.1 System Configuration

Figure 5 shows the overall system configuration for the technology demonstrations.

We set up 8K cameras at the Musashino Forest Sport Plaza, the venue for the Tokyo 2020 Games badminton competitions, and transmitted video images taken during the badminton matches via a 1-Gbit/s network to the broadcast centre. We then used the real-time extraction of objects with arbitrary background[2], a component technology of Kirari!, to selectively extract only the images of players and shuttlecocks from the transmitted 8K images at the broadcast centre. Using highly realistic media synchronization technology (Advanced MMT (MPEG Media Transport))[3], we then synchronously transmitted the extracted images along with multiple images that included video and audio provided by the Olympic Broadcasting Services (OBS) to the remotely located Miraikan museum.

We set up an approximately 100-seater spectator stand and a full-size court equipped with holographic projection equipment at the remote-viewing venue to recreate the actual venue. The extracted images of players and shuttlecocks were holographically displayed by presenting players in their actual positions in front and behind the net using the bird's eye-view sports viewing multi-layer aerial image display technology. As a result, we were able to create a space where the badminton players appeared to be actually present in the remote-viewing venue.

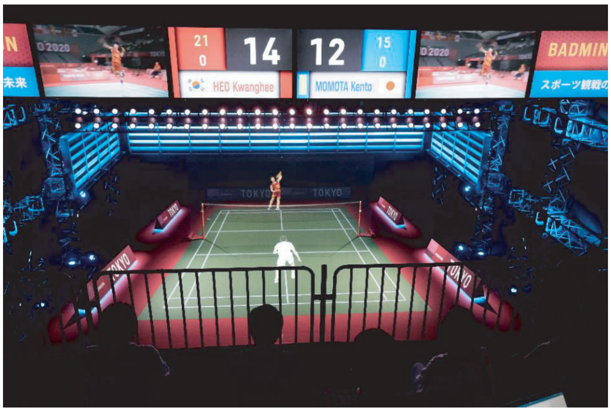


Figure 4 Kirari! Technology Demonstration Experiments of the Tokyo 2020 Games Badminton Competition (© 2021-International Olympic Committee-All Rights Reserved)

3.2 Technical Elements

The following is a list of the technical elements that comprise the system described above.

3.2.1 Real-time Extraction of Objects with Arbitrary Background

The real-time extraction of objects with arbitrary background is a technology for selectively extracting only the specific objects in images, such as players and shuttlecocks, from the videos taken during a competition.

Normally, extraction of only the objects from videos requires using a green or blue background and erasing the background color by chroma keying. This technol-

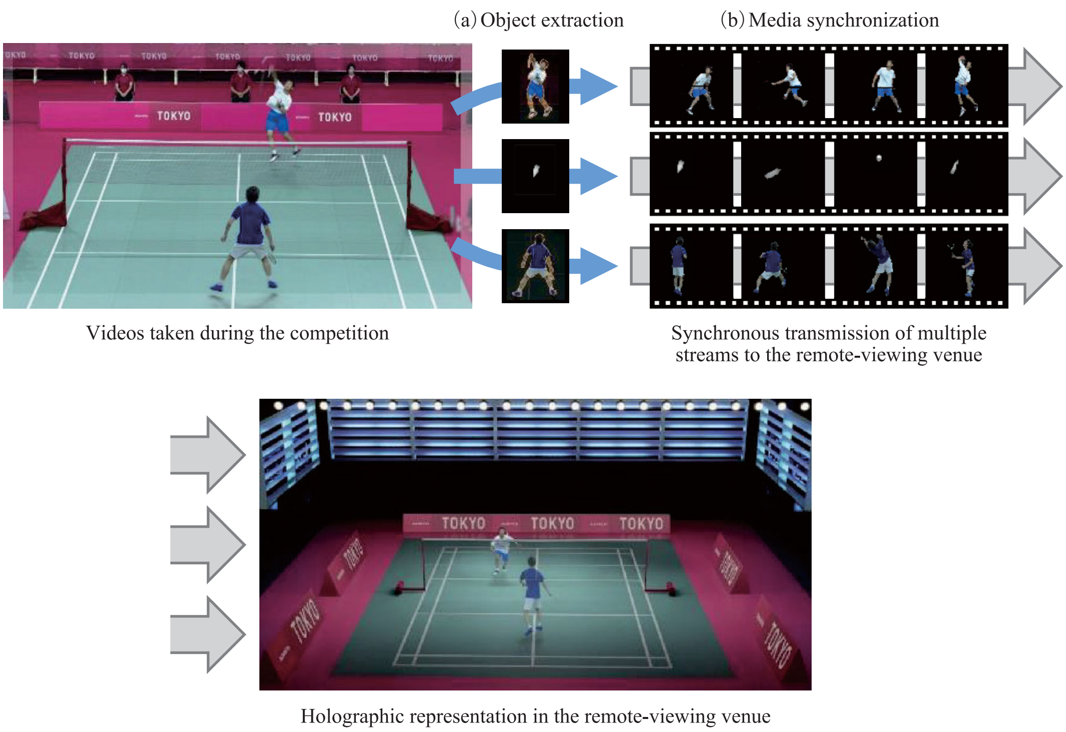


Figure 5 Workflow of Object Extraction and Synchronous Transmission

ogy, however, enables extracting only the objects from the raw images of the competition venue in real time without the need for any special background environment. To apply this technology to the badminton matches, we carried out the following five improvements on our previous system that is for less complicated content.

(1) Individual Extraction of Each Player in the Front and Back Sides of the Court

Separately extracting the front and back players is not possible with the conventional technology. To achieve this, we developed a deep learning model that can simultaneously extract and infer front and back players by inputting depth information simulating the badminton court space. This enabled the extraction of individual players for a game like badminton where players are located separately in the front and back sides of the court. We therefore succeeded in stable and accurate extraction of objects.

(2) Improvement of Output-image Resolution and Frame Rate of Players' Images

We carried out multilayering of the image frames to be processed and leveling of computing resources to support images from 8K, 60-fps cameras. As a result, we were able to produce smooth, high-definition images of players at 930-pixel resolution and 60-fps frame rate.

(3) Stable Extraction of Subtle and Rapid Shuttlecock Movements

We developed extraction algorithms specific for the shuttlecock and succeeded in detecting its exact position and in accurately extracting its various contours. To detect shuttlecock position, we devised a method for learning the shuttlecock's position and movement information by inputting continuous frames in a convolutional neural network (CNN) to eliminate the effects of small objects similar to the shuttlecock appearing in the video (such as seat guide lights). The position of the shuttlecock was determined on the basis of the rough position (heatmap) and correction value (offset) obtained from the CNN. Extraction of the shuttlecock image was carried out by generating the image using a background-subtraction method and filtering with the extracted shuttlecock position and shape and predicted shuttlecock information (position, contour, and motion blur level) in the next frame.

(4) Automatic Correction of Missing Parts in Player Images

When the player at the back of the court overlapped with the net, a black gap zone appeared on the player image. Images were corrected by inferring the missing parts on the basis of the color information above and below the gap zone.

(5) Automatic Generation of Player's Shadow

Reproducing the players' shadows appearing in the actual venue and in the remote-viewing venue enables the creation of more natural player images. We extracted the shadows from the video based on the results of player-image extraction.

3.2.2 Highly Realistic Media Synchronization Technology

The highly realistic media synchronization technology (Advanced MMT) is NTT's proprietary technology developed by extending the media transmission standard MMT and enables the transmitting of various continuous data (streams), such as video, audio, and lighting information, while maintaining their time synchronization. For this project, we synchronously transmitted multiple streams, such as video of the competition taken at the venue, audio, images of players extracted from competition videos, extracted shuttlecock images, video obtained from OBS, etc. and displayed the necessary data on the remoteviewing venue at the appropriate timing to create a highly realistic spatial representation.

3.2.3 Bird's Eye-view Sports Viewing Multi-layer Aerial Image Display Technology

To deliver the experience of watching as if viewers were at the competition venue, we created a space that physically simulated the actual venue (Figure 6).

With our technology, a spectator stand at the same height as the actual competition venue was set up, and the locations of the court, net, two half-mirrors, light-emitting diode displays, and the projectors were optimized to create a realistic appearance of the players in front and behind the net. This enabled the holographically displaying of the two players at their correct positions.

The use of two half-mirrors, however, results in a discontinuous appearance of the shuttlecock, which moves back and forth between the two mirrors. Since this problem is a structural issue arising from having two

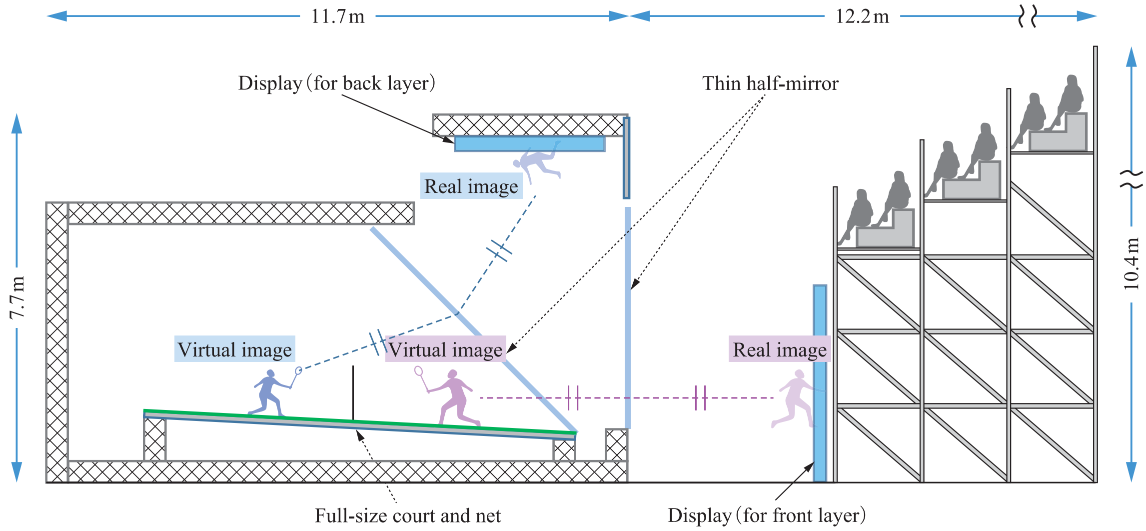


Figure 6 Image of Equipment and Example of Multi-layer Aerial Image Display

physically separated layers, there is no perfect solution to it. To produce a more natural appearance, we repeatedly conducted trial-and-error experiments on the timing for switching the shuttlecock's display layers and came to the conclusion that the best timing for switching is when the player hits the shuttlecock. To determine the best timing for switching, we combined multiple approaches, such as determining the shuttlecock position through analysis of lateral court images, determining the timing when the shuttlecock is hit through detection of hitting sounds from the audio during the match, and human operators watching the game manually switching the display layers. The combination of multiple approaches enabled the real-time switching of shuttlecock-image display positions.

The highly realistic environment with a full-size court and net and life-size holographic projections of players apparently elicited a sensory illusion that the players were actually in front of the participants. This indicates the possibility of achieving an emotional connection at a deeper level than when watching on television. The spectators were able to more easily feel the energy and vibe during the competitions or the excitement of players before the start of the matches, and even the emotions associated with winning or losing. When Kento Momota fell to his knees when he lost in the second round of his match during the group play stage, the realistic sense of his presence in the remote-viewing venue was very striking.

3.3 Results

The following are the technical results of the experiment.

3.3.1 Evaluation of Object Extraction Performance

In terms of object-extraction performance, we succeeded in real-time processing of 8K, 60-fps videos. There are issues that need to be addressed for extraction accuracy such as when images of the two players overlap or extraction for doubles matches. Nevertheless, we succeeded in processing the images without missing parts or errors that could hinder proper viewing of the games, even for players wearing different uniforms for all the six live broadcast experiments. We also confirmed that there were no errors in the correction of the black zones caused by the net and in the addition of shadows. For the shuttlecock images, however, more improvements are necessary due to considerable discontinuity depending on the viewing position.

3.3.2 Transmission Performance Evaluation

In terms of transmission performance, the total end-to-end delay from the competition venue to the remote-viewing venue was less than 2,800 ms. Table 2 lists the lag time details. The MMT transmission delay included network transmission delays of approximately 1 ms between the competition venue and broadcast centre and approximately 0.1 ms between the broadcast centre and remote-viewing venue. Delays in object-extraction processing included buffering that takes fluctuations in processing time, etc., into consideration, but the actual delay in extraction processing excluding the buffering was less than 400 ms.

Since the demonstrations involved one-way transmis-

Table 2　Details of Lag Time

Camera system	≤17 ms（≤1 frame）
MMT transmission	1,600 ms（4K 59.94p BT.709, rate of encode : 40 Mbit/s, FEC : 5%）
Extraction of objects	1,000 ms（including input/output buffering）
Presenting system	～120 ms（4 frames at 30 fps）

sion of images from the competition venue to the remote-viewing venue, an approximately 3-s delay did not cause major problems. However, our future goal is to connect the two venues bi-directionally and deliver the cheers from the audience in the remote-viewing venue back to the competition venue, in which case, significant lag times would become problematic, making it necessary to further reduce processing time.

We were able to confirm that the above performance of the demonstration system was sufficient in delivering a realistic experience of the presence of the athletes in the remote-viewing venue.

4. Conclusion

The planned filming, distribution, and screening of all programs could be performed without significant problems with the dome image distribution. The projection was sufficiently practical in operations since consumer cameras and commercial best-effort lines were used to project images to planetarium systems with entirely different specifications delivered to each museum. Clearly, this approach was practical enough and suggests the possibility of becoming a legacy in the future for the live relay of realistic images of various events.

We applied Kirari! to badminton matches and demonstrated that it can be used to deliver a realistic experience of being at the competition venue. Although there are issues regarding accuracy of extracting player images and in the method for displaying the shuttlecock, we succeeded in demonstrating the possibility of a new sports-viewing experience, beyond what television can provide, for the Olympic Games, an event that attracts attention from all over the world.

Going forward, we will carry out improvements in extraction performance and other technical aspects, as well as deploy the technology to other games and areas, such as for music concerts. We will also study methods of minimizing lag time for bi-directional connection between competition and remote-viewing venues and conduct research and development on Kirari! and other technologies for achieving the Remote World and eventually make future proposals that will surprise the world using communication technologies.

We implemented this project working with Professor Okudo and Researcher Yoshizumi of Wakayama University (the base technical research was supported by JSPS Grant-in-Aid 19K01141), partner companies, people involved in broadcasting, competitions, and conventions, companies and organizations related to each system, as well as local governments and JSCEs nationwide that have gathered citizens and related people and projected the images to them. The projection was supported by many people, including local governments across the country and planetarium facilities such as the National Museum of Emerging Science and Innovation (Miraikan), which held projection events for citizens and other interested parties. We want to express our gratitude to them.

References

（1） K. Nagai, "Evolution History of Broadcast Technology Boosted by Olympic Games," THE JOURNAL OF THE INSTITUTE OF IMAGE INFORMATION AND TELEVISION ENGINEERS, vol. 70, no. 5, pp. 363-371, 2016.

（2） H. Kakinuma, J. Nagao, H. Miyashita, Y. Tonomura, H. Osada, and K. Hidaka, "Real-time Extraction of Objects from Any Background Using Machine Learning," NTT Technology Journal, vol. 30, no. 10, pp. 16-20, 2018.

（3） Y. Tonomura, H. Imanaka, K. Tanaka, T. Morisumi, and K. Suzuki, "Standardization Activities for Immersive Live Experience (ILE)," Journal of the ITU Association of Japan, vol. 47, no. 5, pp. 14-17, 2017.

(Received February 28, 2022 ; Revised March 24, 2022)

KOSHIBA Keiichi

He joined Matsushita Electric Industrial Co., Ltd. in 1989 and was transferred from Panasonic to Tokyo 2020 Organising Committee in 2014. After implementing this project, he left Panasonic then he is verifying and analyzing this project as a visiting fellow at Wakayama University, currently working on the project's legacy as CEO of G1 company Inc.

KINOSHITA Shingo

He received a B.E. from Osaka University in 1991 and M.Sc. with Distinction in technology management from University College London in 2007. He joined NTT in 1991. Until recently, he had been in charge of NTT's Tokyo2020 initiatives, including sports-watching video technology. He has been in his current position since 2021, where he manages R&D on information and communication processing of humans based on human-centered principles.