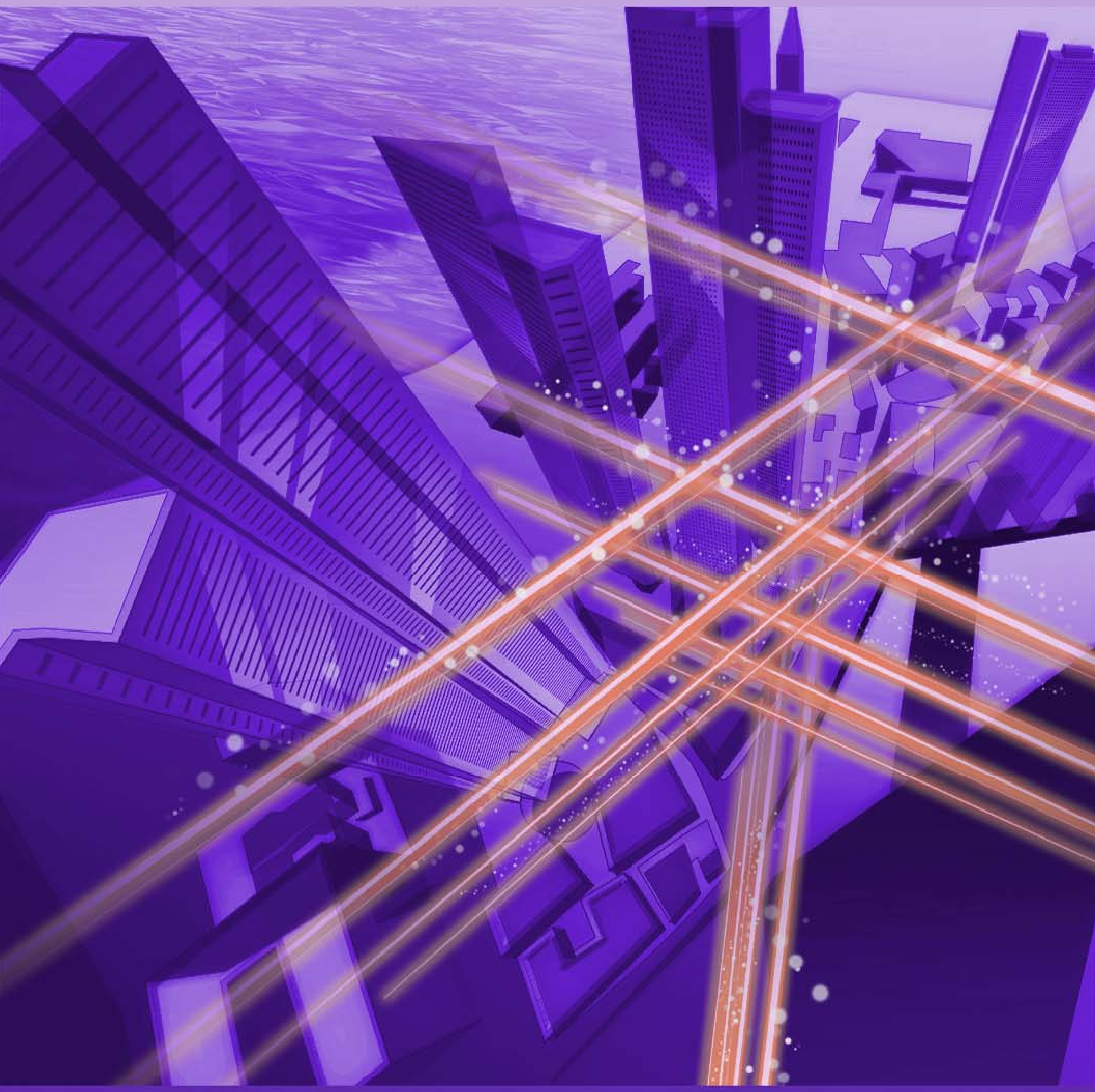
NTT Technical Review

12
2017



December 2017 Vol. 15 No. 12

NTT Technical Review

December 2017 Vol. 15 No. 12

Creation of Immersive UX Services

Ryuji Kubozono, Akihito Akutsu, Norihiko Matsuura, Kenichi Minami, and Akira Ono

Abstract

The NTT Group is researching and developing communication technology and media processing technology amid growing expectations for the creation of services that provide users with a highly realistic experience of events such as sports matches and live entertainment shows. We are also working with various corporate partners to conduct feasibility studies and service trials aimed at creating new value. This article describes the direction of our research and development aimed at evolving the key technologies needed to implement immersive user experience (UX) services, and presents an overview of our efforts to create entirely new immersive UX services.

Keywords: high reality, public viewing, live entertainment

1. Current state of media services

In the television (TV) corners of large consumer electronics retailers nowadays, most of the TV sets are 4K-compatible. With the ongoing advances in the resolution of consumer video cameras and digital cameras, we are now starting to see 4K TV broadcast services, and even trials of 8K broadcasting. The move to higher resolution is also driving development in the audio market, and it is expected that this trend will continue in the future as the world's media-related services and products migrate towards higher definition and higher resolution.

In addition to this market trend, 2016 has also been called the first year of the virtual reality (VR) era due to the emergence of various products and services using VR and/or augmented reality (AR), especially in the game/amusement sector. In addition to specialist VR/AR equipment, the market is now awash with inexpensive head-mounted displays (HMDs) and smartphone-based VR/AR applications, and we are starting to see a wide variety of content being aimed at these terminals. For example, Pokémon GO was a major success around the world and became the first such application to really capture the public's imagination.

Meanwhile, amusement parks and movie theaters aim to provide a more realistic viewing environment, and in addition to making advances in three-dimensional (3D) video and higher-quality audio with more channels, systems such as MX4D™ [1] and 4DX® [2] are also incorporating other elements besides audio and video in order to stimulate the other senses with, for example, vibrations, water sprays, mists, or aromas. When combined with audio and video being displayed in front of the audience, these services provide an even greater sense of realism. Services of this sort have been a popular feature of amusement parks for many years, and efforts are being made to advance this technology into movie theaters so that a trip to the movies will become more of an experience than simply watching a film.

There are also facilities that achieve a heightened sense of realism with audio and video by projecting video productions onto 360° screens or dome-shaped screens, or by covering the audience's field of view with three screens (in front and on both sides) [3], and we are expecting to see more movie theaters and other entertainment facilities introducing immersive environments in the future.

2. Current state of the entertainment sector

In the field of sports, attending games and competitions has been a popular past time for many years. However, not everyone is able to attend such events at the venues where they are occurring. Consequently, public viewings are becoming more common in Japan for games and competitions that attract a lot of interest, especially when Japanese teams are competing at international events. Various forms of public viewings are held in places such as sports stadiums and public halls so that the action can be enjoyed by fans who were unable to travel to the actual venue.

Meanwhile, movie theaters are showing non-movie content (ODS: other digital source) during intervals in movie screenings and are promoting measures aimed at attracting new customers. It is predicted that by 2020 this will have grown into a ¥63.3 billion market (including ¥31.8 billion for live broadcasts) [4]. Sports-related public viewings are often held free of charge, but we are starting to see viewings that charge an entrance fee. As progress is made in the resolution of issues relating to content rights, there will probably be a growing number of monetized cases.

In the area of music entertainment, the music industry has known for many years that sales of packaged music such as compact discs are set to decline and is shifting towards holding live events where music fans can get a once-in-a-lifetime experience even if the artist is playing the same set list.

Compared with western countries, Japan still has relatively strong sales of packaged music, but the number of live performances is rising, and it is no exaggeration to say that this is part of a major worldwide trend [5]. Although the live music market is continuing to grow, many concerts offer new experiences through the use of elaborate stage productions and gimmicks to ensure that the appetite of audiences for repeated viewings is undiminished. The latest technologies and production methods are being introduced for this purpose, and we can expect this trend to continue into the future.

Digital technology has also recently been used in various ways for productions other than music concerts. Many different types of works based on comics, films, and role-playing games are showcased on stage with the latest digital technology, and it is possible to reproduce the world view of the original creator to a high degree and to enhance the audience experience. The latest digital technology has also been incorporated into traditional stage performances. In a production of *The Tempest* at the Shakespeare Theatre in London [6], motion capture and projection mapping technologies were used to create mysterious special effects even within a traditional theatrical production. In Japan, performances that combine traditional *kabuki* theater with the latest projection mapping techniques have attracted considerable interest.

3. NTT's vision of immersive UX services

At NTT, in light of the abovementioned market trends, we aim to create immersive user experience (UX) services to provide audiences with new experiences. To realize this goal, we aim to develop various applications, including *ultra-immersive public viewing*, where people can experience the strength and speed of the world's top sports players even when they are far away from the actual sporting venue, *ultra-immersive live viewing*, where people at a remote location can enjoy traditional performing arts in the same way as people in the theater or hall where the performance is actually taking place, *enhanced live performances*, where people can experience something newer than an ordinary live performance, and *stadium solutions*, where people can experience seeing a sports match in person, but with a great deal of added fun.

3.1 Ultra-immersive public viewing and live viewing

The aim of ultra-immersive public viewing and live viewing (**Fig. 1**) is to implement a world where sports events or live performances can be transmitted in their entirety to remote locations, where people can experience the events as if they were there themselves. During large international sporting events, we already have public viewings using video footage displayed on large screens in stadiums throughout Japan. However, this is not a highly immersive experience. Instead, we aim to use a combination of display control technology that enables people to view the event as if they were in the same venue as the sports players or performers, presentation control technology that enables the audience to feel as if they had entered an actual sports venue, and audio technology that enables them to experience new sounds. Furthermore, we aim to provide a new viewing experience by detecting and tracking the subjects appearing in the video images and linking them to various kinds of metadata, resulting in a service that would be impossible to implement with TV broadcasting.

3.2 Enhanced live performances

We are looking at ways of providing greater emotional impact in traditional performing arts (**Fig. 2**). Since 2016, NTT has been providing technology to the Cho-Kabuki series of events in partnership with

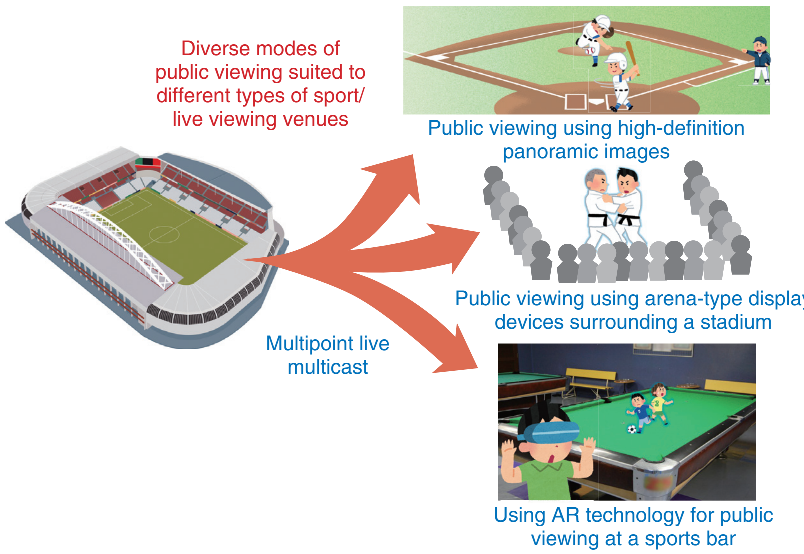


Fig. 1. Ultra-immersive public viewing concept.

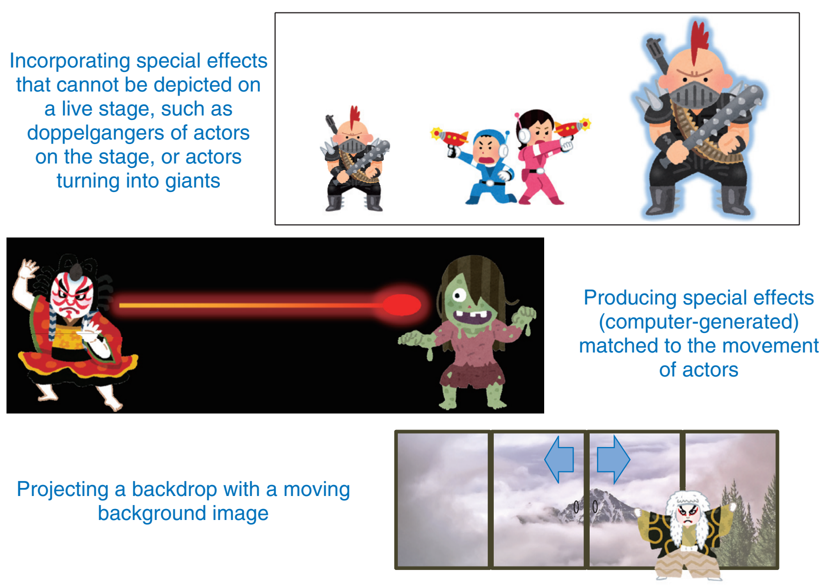


Fig. 2. Enhanced live performances.

Dwango Co., Ltd. and Shochiku Co., Ltd. in order to provide new performance events by combining the traditional art of kabuki with modern technology. We have also been working at identifying potential applications for this technology and extracting new technical issues.

3.3 Stadium solutions

A stadium solution (**Fig. 3**) is a way of entertaining spectators at a sports event before and after the game, and increasing their enjoyment of the game itself so that people will want to visit the stadium even on days when no games are taking place. In this way, we aim to revitalize stadiums and their surroundings. For example, NTT is working on the creation of a showcase

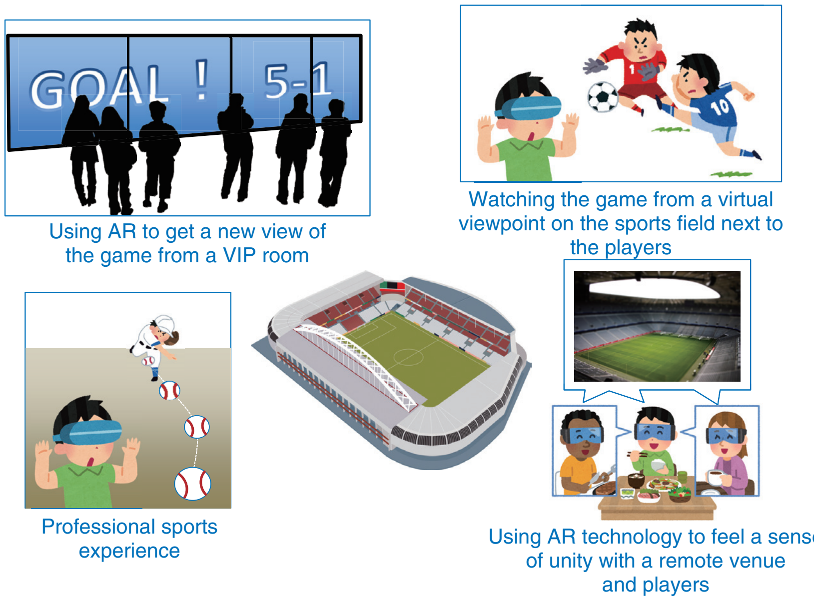


Fig. 3. Stadium solution concept.

where people can use VR technology to experience what it would be like to face balls thrown or hit towards them by professional baseball and tennis players, and is evaluating its feasibility as a VR experience.

4. Main technological components of an immersive UX service

We can provide new ways to enjoy sports by allowing people to see and hear things that they would not normally be able to experience. This opens the door to a completely new way of looking at sporting events. One technology that can be used to implement such a system is called free-viewpoint video synthesis technology [7]. This enables people wearing HMDs to watch sports from the viewpoint of a restricted part of the stadium that they cannot enter. The use of high-speed cameras makes it possible to show viewers the exact course traveled by the ball and players, and high frame rate video encoding technology can be used to compress the video pictures with high definition and a high frame rate to show subtle movements and differences in dynamism [8].

In addition, the use of surround video stitching and synchronous transmission technology makes it possible to realize a super high definition panoramic video picture covering the entire field of view instead of simply presenting the pictures on a large TV screen as in previous public viewings [9]. The use of wave field synthesis technology can position a sound image at the location of the subjects shown in the video, or at a position away from the front of the display screen, enabling multiple audience members to experience relayed sound without having to provide separate headphones or other audio devices for each individual [10].

Furthermore, the use of goggle-free 3D video screen technology that enables the viewing of stereoscopic images without having to wear 3D goggles will make it possible to provide a completely new kind of highly immersive experience [11]. To convert images into 3D, it is necessary to extract the subject of interest from video pictures captured from many different angles by means of arbitrary background real-time object extraction technology [12]. For sports and entertainment events, we must consider how accurately this can be done in real time against backgrounds that contain moving objects.

In addition, by displaying various additional information together with the audiovisual content, we can enhance people's enjoyment of sports and entertainment events and help them to learn about the players and performers. This increases the appeal of relayed broadcast events. An important component of such systems would be moving object detection technology that can detect the positions and postures of players on a sports field and of performers on a stage [13].

We expect that this will be the subject of active research and development in the future.

5. Future prospects

In addition to the technologies introduced here, the NTT Group is pursuing open innovation in various fields to create new kinds of immersive UX services based on diverse media processing and communication technologies. As a result, although we are currently pursuing many technologies with a view to business development, there are still many others that remain at the level of feasibility studies. Going forward, we will continue to pursue open innovation with various players in order to establish media processing and communication technologies and create services that can be used to address social issues and stimulate local growth.

References

[1] Website of Toho Cinemas' MX4D (in Japanese), https://www.tohotheater.jp/service/mx4d/

[2] Website of United Cinemas' 4DX (in Japanese), http://www.unitedcinemas.jp/4dx/

[3] Website of United Cinemas' ScreenX (in Japanese), http://www.unitedcinemas.jp/screenx/

[4] Press release issued by GEM Standard on February 3, 2016 (in Japanese). https://gem-standard.com/information/100

[5] PIA Research Institute, "2016 White Paper on Live Entertainment," 2016 (in Japanese).

[6] Website of d3 Technologies, "The Tempest," http://www.d3technologies.com/en/showcases/theatre/the-tempest/

[7] K. Okami, K. Takeuchi, M. Isogai, and H. Kimata, "Free-viewpoint Video Synthesis Technology for a New Video Viewing Experience," NTT Technical Review, Vol. 15, No. 12, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201712fa2.html

[8] Y. Omori, T. Onishi, H. Iwasaki, and A. Shimizu, "A 120 fps High Frame Rate Real-time Video Encoder," NTT Technical Review, Vol. 15, No. 12, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201712fa3.html

[9] 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, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201712fa4.html

[10] K. Tsutsumi and H. Takada, "Powerful Sound Effects at Audience Seats by Wave Field Synthesis," NTT Technical Review, Vol. 15, No. 12, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201712fa5.html

[11] M. Makiguchi and H. Takada, "Smooth Motion Parallax Glassless 3D Screen System that Uses Few Projectors," NTT Technical Review, Vol. 15, No. 12, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201712fa6.html

[12] H. Nagata, H. Miyashita, H. Kakinuma, and M. Yamaguchi, "Real-time Extraction of Objects with Arbitrary Backgrounds," NTT Technical Review, Vol. 15, No. 12, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201712fa7.html

[13] T. Tokunaga, Y. Tonomura, J. Shimamura, F. Hamamura, and H. Kakimoto, "Real-time Moving Object Detection Technology and Trial of Stone Location Delivery at a Curling Venue," NTT Technical Review, Vol. 15, No. 12, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201712fa8.html

Trademark notes

Ryuji Kubozono
Vice President, NTT Service Evolution Laboratories.

He received a B.S. and M.S. in physics from Kagoshima University in 1987 and 1989. He joined NTT Human Interface Laboratories in 1989. He was with NTT WEST from 1999 to 2003 and from 2006 to 2008, and with NTT Smartconnect from 2012 to 2016.

Akihito Akutsu
Executive Research Engineer, Proactive Navigation Project, NTT Service Evolution Laboratories.

He received an M.E. in engineering from Chiba University in 1990 and a Ph.D. in natural science and technology from Kanazawa University in 2001. Since joining NTT in 1990, he has been engaged in research and development (R&D) of video indexing technology based on image/video processing, and man-machine interface architecture design. From 2003 to 2006, he was with NTT EAST, where he was involved in managing a joint venture between NTT EAST and Japanese broadcasters. In 2008, he was appointed Director of NTT Cyber Solutions Laboratories (now NTT Service Evolution Laboratories), where he worked on an R&D project focused on broadband and broadcast services. In October 2013, he was appointed Executive Producer of 4K/8K HEVC (High Efficiency Video Coding) at NTT Media Intelligence Laboratories. He received the Young Engineer Award and Best Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 1993 and 2000, respectively. He is a member of IEICE.

Norihiko Matsuura
Executive Research Engineer, NTT Media Intelligence Laboratories.

He joined NTT in 1994 and engaged in R&D of a 3D shared space system for communication via avatars in a virtual environment. He is currently in charge of the Visual Media Project, where he is conducting R&D in areas such as video encoding/decoding and artificial intelligence technologies for visual media.

Kenichi Minami
Executive Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E. in electronic engineering and an M.S. in biomedical engineering from Keio University, Kanagawa in 1991 and 1993. He received an MBA from Thunderbird, Global School of Management, Arizona, USA, in 2002. He has been engaged in R&D management in the development of “Kirari!” immersive telepresence technology since 2016. During 2012–2014, he was responsible for the development of mobile application services at NTT DOCOMO. His research interests include image and audio processing, user interfaces, and telepresence technologies. He is a member of IEICE.

Akira Ono
Senior Research Engineer, Supervisor, Natural Communication Project, NTT Service Evolution Laboratories.

He received an M.E. in computer engineering from Waseda University, Tokyo in 1992. He joined NTT in 1992 and was involved in R&D of video communication systems. From 1999 to 2010, he worked at NTT Communications, focusing on network engineering and creating consumer services. He moved to NTT Cyber Solution Laboratories (now, NTT Service Evolution Laboratories) in 2010. Since 2015, he has been studying the immersive telepresence technology called “Kirari!”.

Free-viewpoint Video Synthesis Technology for a New Video Viewing Experience

Kazuki Okami, Kouta Takeuchi, Megumi Isogai, and Hideaki Kimata

Abstract

There is a growing desire among viewers of sports programs to watch videos of matches with an even higher sense of presence. Free-viewpoint video synthesis has been attracting attention as one scheme for achieving a video viewing experience with an ultrahigh sense of presence. This technology reconstructs the match space targeted for viewing and provides video from the viewpoint of the user's choosing. This article describes the technologies for free-viewpoint video synthesis developed by the NTT laboratories and introduces application examples.

Keywords: free-viewpoint video, video viewing, image processing

1. Introduction

Free-viewpoint video synthesis enables the space targeted for a sports competition or event to be viewed from an arbitrary viewpoint as desired by the user. The viewpoint of synthesized video is not restricted by the arrangement of the cameras used for shooting. That is to say, the technology achieves video viewing from a *free* viewpoint; it does not simply switch cameras or interpolate between them. For example, in a soccer match, free-viewpoint video synthesis can achieve video viewing even from positions where no cameras are installed such as a viewpoint alongside a player standing in the field or a viewpoint tracking the ball. In short, free-viewpoint video synthesis can provide the user with a viewing experience that exceeds the experience of actually being at the stadium.

2. Flow of free-viewpoint video synthesis and technology trends

Various methods of free-viewpoint video synthesis have been developed [1]. We introduce a geometry-based technique that we have been working on. The process begins with capturing the scene desired for viewing and restoring three-dimensional shape and texture (image) information. Then, at viewing time, the system renders an image of such three-dimensional shapes and textures watched from the desired viewpoint.

To give some background, we describe Eye Vision, a well-known viewing system used in the 2001 Super Bowl telecast in the United States. The technology at that time involved shooting the target scene from various positions and directions by installing a large number of cameras on a scale of several tens of units and estimating three-dimensional information by performing image processing such as stereo matching [2] with the multi-view images obtained.

Today, however, high-definition cameras are becoming commonplace, so shooting a scene with many cameras and carrying out image processing using many multi-view images is considered to be prohibitive in time and cost.

However, a method developed recently can use both color images taken with ordinary cameras and depth images obtained with depth sensors [3]. Here, a depth

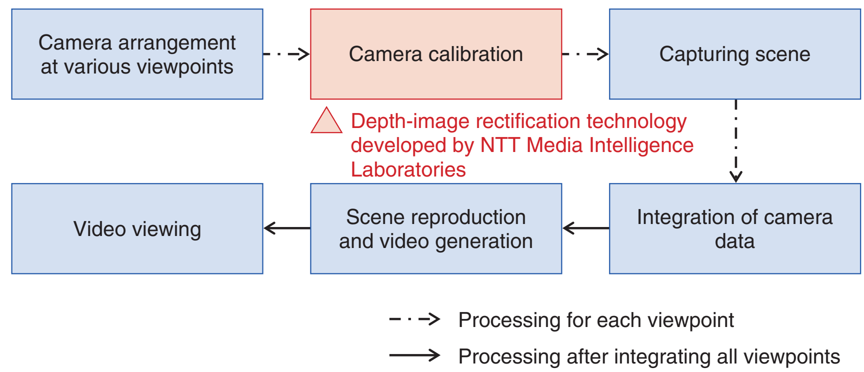


Fig. 1. Free-viewpoint video synthesis technology.

image represents the distance to an object measured by a depth sensor in the form of an image. It can directly represent the three-dimensional shape of an object from the shooting direction, which means that free-viewpoint video synthesis can be achieved with fewer cameras than the conventional technique of estimating three-dimensional information from multi-view images.

In addition, low-cost image-pickup devices equipped with a depth sensor have recently become available such as the Microsoft Kinect (Kinect). Such devices can be used to construct a low-cost system for shooting and synthesizing video from arbitrary viewpoints.

As described above, achieving free-viewpoint video synthesis in a wide-area space in a stadium or other venue is a near-future objective, but the first step here will be to use a depth sensor such as Kinect to establish free-viewpoint video synthesis technology targeting the relatively small space surrounding a person. The rest of this article describes the technology being researched and developed at NTT Media Intelligence Laboratories, presents a demonstration system, and introduces application examples.

3. Depth-image rectification technology for free-viewpoint video synthesis using depth sensors

We introduce here a small-scale demonstration system we constructed for researching free-viewpoint video synthesis using four Kinect units, each with a combined camera and depth sensor. The system flow and the shooting environment are respectively shown in **Figs. 1** and **2**. This system uses only a few compact Kinect devices as shooting cameras, making it possible to synthesize free-viewpoint video with high portability and at low cost. However, the mechanism and performance of this depth sensor and other factors limit the subject to a small space of about three square meters.

The four Kinect devices are each connected to a client personal computer (client PC). The system captures both color and depth images of the subject in the space and sends those images to the synthesis PC via a network switch. Then, by creating a correspondence between each camera and image, the system reproduces the captured subject as point-cloud data (a collection of points) or as mesh data that smoothly connect the point-cloud data.

We explain here the key elements of this technology that we are researching and developing. To obtain point-cloud data of the subject from color and depth images, it is necessary to calculate beforehand camera parameters such as the relative positions, orientations, and lens-related parameters of the cameras for both color images and depth images. This parameter-calculation process is called calibration. Specifically, there are two types of calibration for each camera: external and internal. External calibration is a process that determines the position of a camera. It must be performed to integrate the depth images received from each Kinect unit at their correct positions.

Internal calibration, meanwhile, corrects for lens distortion in the images captured by each camera. This calibration must be performed to obtain accurate images and point-cloud data.

Calibration using only color images has been used for a long time, and parameters have been obtained by performing the above external and internal calibration using a technique [4] featuring a structure with a fixed pattern (such as a black and white checkerboard).

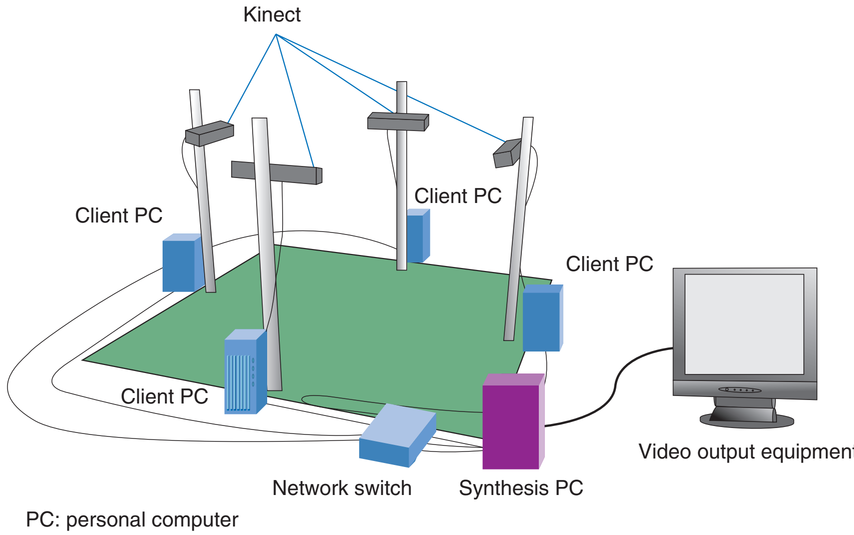


Fig. 2. Shooting environment.

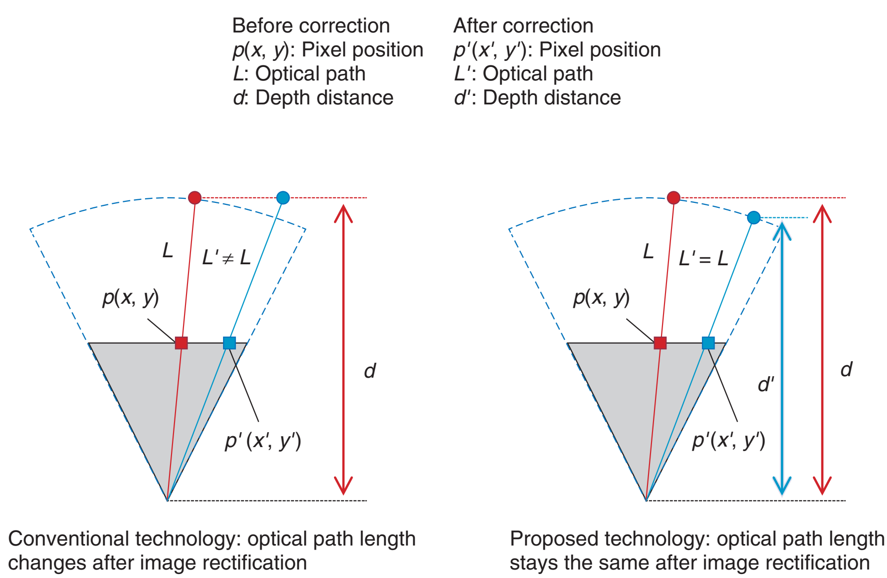


Fig. 3. Conceptual diagram of depth-image correction technology.

However, the depth images cannot be sufficiently corrected by conventional internal calibration, which means that accurate depth images cannot be obtained. This problem originates in the technique used to obtain the depth images. A Kinect device irradiates the target object with infrared light, obtains the optical path length by measuring the time taken for the light to be reflected and to return, and calculates the perpendicular distance to the object from the optical path length as the depth value. Consequently, if we use only camera lens distortion taken into account by conventional internal calibration, the result is a depth value calculated from an optical path different from the correct one, as shown in **Fig. 3**.

To solve this problem, we proposed a technique for depth-image rectification [5] that correctly calculates depth values after lens rectification by using mapping positions before and after lens distortion rectification

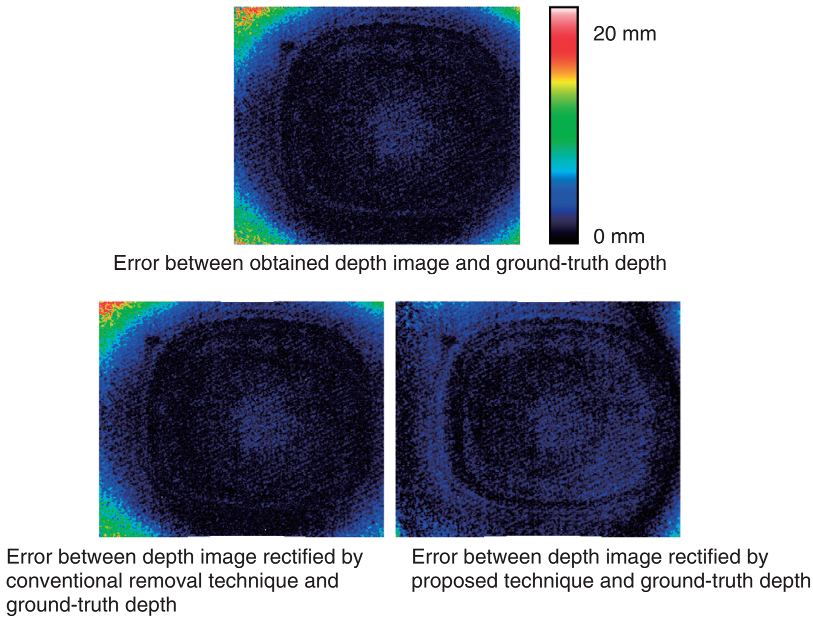


Fig. 4. Comparisons of error with ground-truth depth of depth image.

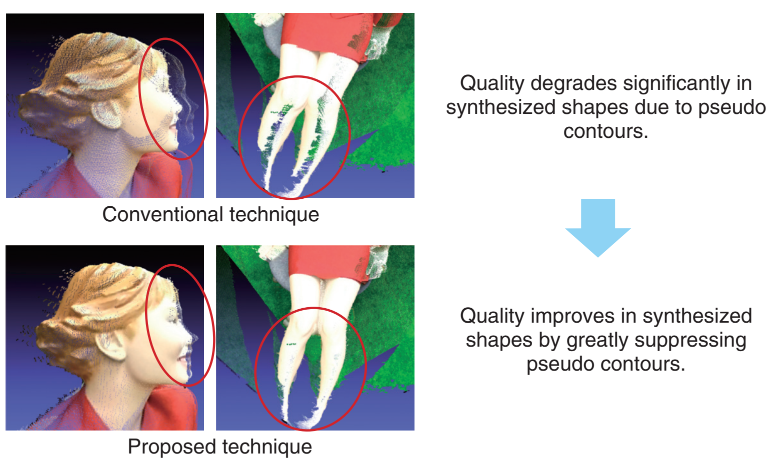


Fig. 5. Comparison of results applying depth-image rectification techniques.

and constraints that prevent the obtained optical-path values from changing. This technique makes it possible to obtain depth images with less error than that using the conventional calibration technique (**Fig. 4**).

A comparison of reconstruction results using the conventional technique and our proposed technique is shown in **Fig. 5**. The conventional technique results in quality degradation due to the appearance of pseudo contours of a subject. In contrast, the proposed technique greatly suppresses pseudo contours, thereby significantly contributing to improved quality in reproducing the three-dimensional shape of the subject.

4. Applications of free-viewpoint video synthesis in small spaces

We are considering a real-time system and an offline one for free-viewpoint video synthesis application in small spaces. The real-time system displays

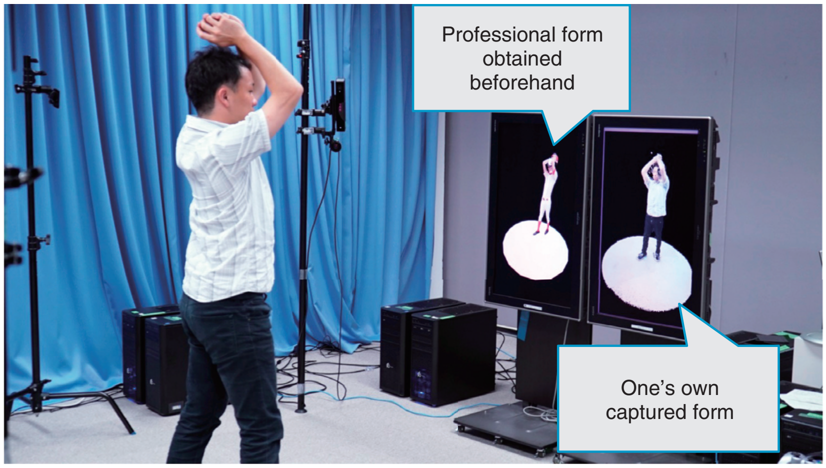


Fig. 6. Application to self-check one's form in real time.

the images obtained from the Kinect units as point-cloud data in real time. For example, we can imagine an application that would enable an athlete or dancer to check his or her form in real time. As shown in **Fig. 6**, lining up one's own actions in the present with those of a professional athlete obtained beforehand makes it possible to see where one might have poor form (e.g., dropping of the elbow compared with that of a professional baseball pitcher). Moreover, as this application would perform three-dimensional reconstruction different from simple images or video, it would make it possible to compare postures from an arbitrary viewpoint that would not normally be available. An application of this type should provide a more efficient means of correcting defects in one's athletic or artistic form.

The off-line system, on the other hand, would first integrate the images obtained from the Kinects as point-cloud data on the synthesis PC and then perform mesh processing to create surface data from the point-cloud data. Then the system would render the surface data instead of simply rendering the point-cloud data. This processing would enable the rendering of a three-dimensional model of even higher quality and the reconstruction of a high-density, smooth three-dimensional model of the subject (**Fig. 7**).

Fig. 7. Three-dimensional reproduction using proposed technology.

At present, however, the computational cost is too high to achieve real-time rendering. The off-line application could therefore be used to view three-dimensional shapes of a subject at a leisurely pace after shooting has been completed. One example of such an application would be self-checking of one's form just as in the real-time application. Here, however, instead of checking one's form instantly, the application could be used to save one's own model on a device such as a smartphone and review it whenever and wherever one likes.

It would also be possible to place a three-dimensional model of oneself in a virtual space and use a head-mounted display (HMD) to check one's life-size form. Another application differing from the real-time applications is to use it as a three-dimensional photo. This application could be used in many enjoyable ways such as lining up three-dimensional models of oneself and one's friends, viewing them from any direction on a smartphone or HMD, and adding lighting effects.

5. Future development

As the first step to achieving free-viewpoint video synthesis, we researched and developed technologies for synthesizing small spaces with an emphasis on portability while envisioning applications for trouble-free capturing and rendering of the space surrounding a person. In this article, we outlined free-viewpoint video synthesis technology and a demonstration system and introduced promising application examples. As the next step, we plan to target larger spaces such as stadiums and provide a video viewing experience that can make the viewer feel like a user standing in any location even if no cameras are installed at that point. Achieving such an application will require free-viewpoint video synthesis oriented to larger spaces, and to this end, we are researching and developing technologies for capturing and synthesizing objects in such an environment.

Furthermore, while the free-viewpoint video synthesis system that we have so far researched and developed performs synthesis processing on site, that is, at the capturing location, we are now developing techniques for transferring data to the cloud over the network and performing synthesis processing on the cloud. This will enable viewing of free-viewpoint video content from remote locations and provide a more enjoyable viewing experience.

References

[1] A. Smolic, "3D Video and Free Viewpoint Video—From Capture to Display," Pattern Recognit., Vol. 44, No. 9, pp. 1958–1968, 2011.

[2] C. Zitnick and T. Kanade, "A Cooperative Algorithm for Stereo Matching and Occlusion Detection," IEEE Trans. Pattern Anal., Vol. 22, No. 7, pp. 675–684, 2000.

[3] D. Alexiadis, D. Zarpalas, and P. Daras, "Fast and Smooth 3D Reconstruction Using Multiple RGB-Depth Sensors," Proc. of the IEEE Visual Communications and Image Processing Conference 2014, pp. 173–176, Valletta, Malta, Dec. 2014.

[4] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Trans. Pattern Anal., Vol. 22, No. 11, pp. 1330–1334, 2000.

[5] A. Louie, K. Takeuchi, N. Ito, and A. Kojima, "Multiple Plane View Camera Calibration for RGB-D Sensor Rectification," IEICE Tech. Rep., Vol. 115, No. 350, IE2015-93, pp. 81–86, 2015.

Trademark notes

Kazuki Okami

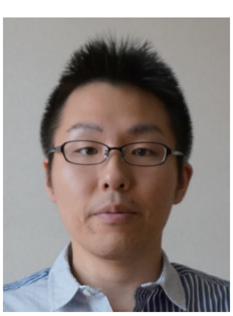
Researcher, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. from Waseda University, Tokyo, in 2012 and 2014. He joined NTT Media Intelligence Laboratories in 2014, where he has been engaged in research and development of free-viewpoint video synthesis.

Megumi Isogai

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

She received a B.E., M.E., and Ph.D. in communication network engineering from Okayama University in 2004, 2005, and 2010. She joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 2006 and has been studying three-dimensional video processing and high-reality communication technology.



Kouta Takeuchi

Engineer, NTT Plala Inc.

He received a B.E. and M.E. in engineering from Nagoya University, Aichi, in 2009 and 2011. He joined NTT in 2011 and conducted research on free-viewpoint video synthesis and super-resolution of a light field camera. Since July 2017, he has been with NTT Plala and working on a set-top box development project.

Hideaki Kimata

Senior Research Engineer, Supervisor, High-Reality Visual Communication Group Leader, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in applied physics in 1993 and 1995 and a Ph.D. in electrical engineering in 2006 from Nagoya University, Aichi. He joined NTT in 1995 and has been researching and developing a video coding algorithm, visual communication systems, and machine learning. He is a Chief Examiner of the Information Processing Society of Japan (IPSJ) Special Interest Group on Audio Visual and Multimedia information processing.

A 120 fps High Frame Rate Real-time Video Encoder

Yuya Omori, Takayuki Onishi, Hiroe Iwasaki, and Atsushi Shimizu

Abstract

This article describes a real-time HEVC (High Efficiency Video Coding) encoder operating at 120 frames per second (fps) that is designed to achieve higher frame rate video services. The encoder achieves 4K/120 fps video encoding in real time through the synchronized operation of multiple 2K/120 fps encoders working in parallel. This encoder also makes it possible to achieve temporal scalable coding and transmission with upward compatibility for existing 60 fps based systems. This temporal scalability is expected to contribute to rapid expansion of the high frame rate video service field. The proposed encoder systems will also open the door to next-generation high frame rate ultra-high definition television services.

Keywords: high frame rate, encoder, hardware

1. Introduction

The latest video coding standard, H.265/HEVC (High Efficiency Video Coding), achieves double the coding efficiency of H.264/AVC (Advanced Video Coding), making it possible to provide higher definition video services economically. In recent years, 4K video broadcasting and distribution have become increasingly widespread, and realistic image representations are becoming increasingly popular. Such representations demand not only high spatial resolution but also many other factors (**Fig. 1**).

A high frame rate (HFR) improves the moving picture quality and is essential for creating more realistic image representations [1]. The next-generation television system specified in Recommendation ITU-R*1 BT.2020—Parameter values for ultra-high definition television (UHDTV) systems for production and international programme exchange—supports HFR formats up to 120 fps. For the spread of HFR video services, temporal scalable coding with upward compatibility for current 60 fps based video services is also necessary for encoders. Several HEVC real-time encoders have been developed to enable HEVC encoding of over-4K images. However, they are only capable of encoding images up to 60 fps due to the increase in computational complexity and the need for temporal scalable functionality.

This article presents our new 4K/120 fps HEVC encoder, which achieves HFR 120 fps real-time encoding and 120/60 fps temporal scalable functionality by exploiting our 4K/60 fps HEVC codec large-scale integrated circuit (LSI) called NARA, which stands for Next-generation Encoder Architecture for Real-time HEVC Applications [2].

When the flexible customizable software architecture of the NARA LSI is utilized, merely modifying the custom functions layer of the NARA software architecture makes it possible to achieve temporal scalability such as upward compatible reference picture structures and dual-stream bit-rate control functions. An encoder equipped with one NARA LSI has 2K/120 fps encoding capability, and synchronized operation of multiple encoders working in parallel also achieves scalability to larger 4K/120 fps input images. These scalable functionalities of the proposed encoder will contribute to the development of

*1 ITU-R: International Telecommunication Union - Radiocommunication Sector

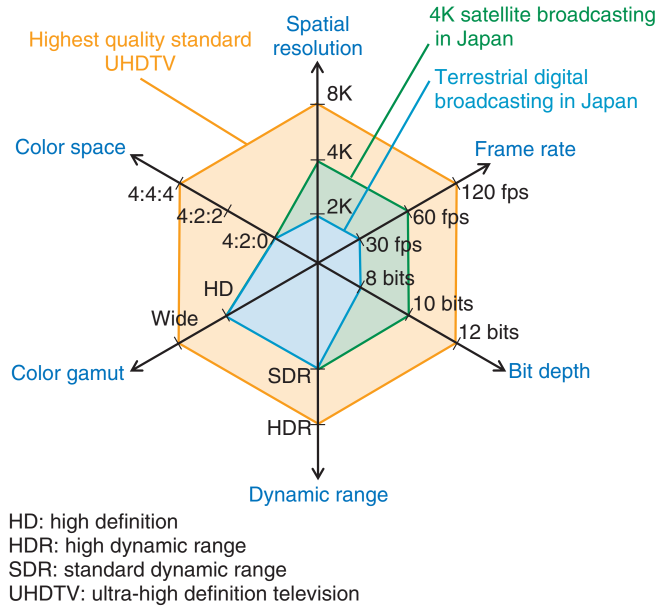


Fig. 1. Factors required for high quality video.

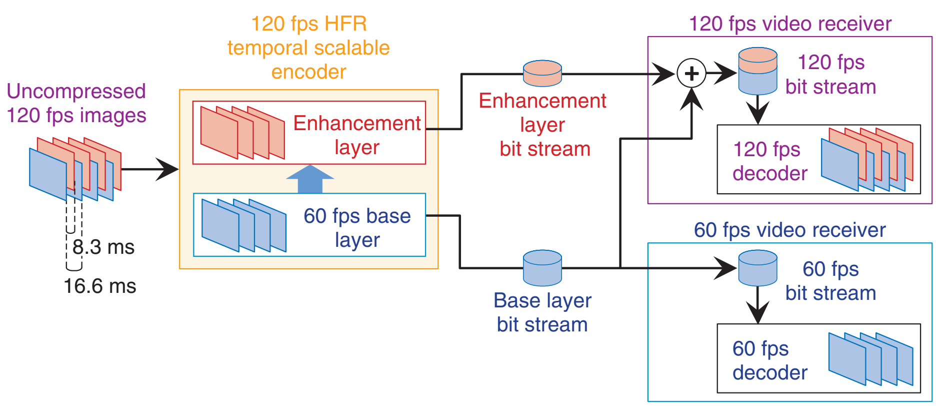


Fig. 2. HFR temporal scalable video distribution.

new broadcasting and distribution systems for the upcoming UHDTV services.

2. Encoder system architecture for HFR temporal scalability

The Association of Radio Industries and Businesses (ARIB), an incorporated association promoting the practical application and dissemination of radio systems in Japan, regulates 120/60 fps temporal scalable formats as an HFR scalable coding standard based on HEVC [3]. The ARIB standard specifies that 120 fps HFR bit streams must have temporal scalability, and that encoded picture data are distributed into a 60 fps base layer stream and an enhancement layer stream, as illustrated in **Fig. 2**. Dual-stream bit-rate control must be performed for both the base layer and enhancement layer streams to ensure constant bit-rate encoding and distribution for both 60 fps and 120 fps decoders. In addition, base layer images and enhancement layer images should be periodically received and decoded one by one alternately to prevent deviation of the decoding time for 60 fps decoders. This limitation of the decoding time leads to changes in the

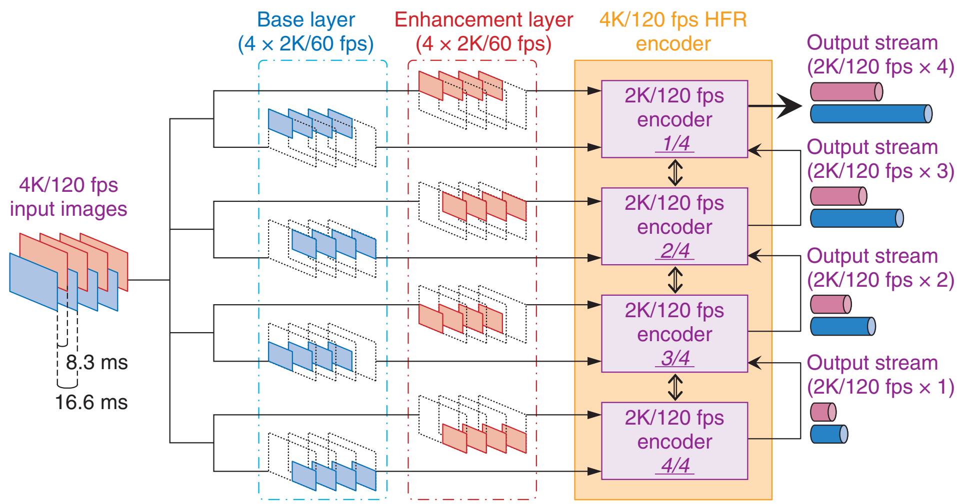


Fig. 3. System configuration.

reference structure of inter-frame prediction.

The temporal scalable encoding function is added to the existing NARA LSI by utilizing the flexible customizable software architecture of a NARA LSI with large motion search capability. The LSI's software architecture consists of three hierarchical layers; the top function layer is the software for handling fundamental HEVC functions and user functions. This software hierarchy not only solves the complexity of tediously controlling the hardware using a low-level interface and the difficulty of handling HEVC common basic functions, but also provides a simple programming interface as custom functions of the top layer. Temporal scalability essentially requires complicated modifications of the LSI's encoding method. However, because of this software architecture, the dual-stream bit-rate control and the reference structure modification are achieved in some of the custom functions, and thus, they can be easily customized with higher level programming.

3. System configuration

The system configuration of the 4K/120 fps encoder we have developed is shown in **Fig. 3**. It consists of four 2K/120 fps encoders. Each 2K/120 fps encoder includes one NARA LSI and handles the encoding of one of the 2K images, which is squarely divided from the original 4K input images. The 2K/120 fps encoder receives a 2K/120 fps input video as two sets of 2K/60 fps video sequences, which correspond to the base layer and the enhancement layer, by using a multi-channel input functionality of the NARA LSI. The encoder rearranges the two 2K/60 fps sequences to form one 2K/120 fps image, and all 2K/120 fps encoders operate cooperatively by exchanging synchronization signals in order to share the common clocking and time stamp values to form one 4K/120 fps encoder.

The output stream is constructed in the MPEG-2 Transport Stream (TS)*2 format so that the base layer stream and the enhancement layer stream have different video packet identifiers. Here, the NARA LSI's multi-channel video output function is effectively utilized. Four 2K/120 fps transport streams can be transmitted in parallel, and they can also be transmitted with one multiplexed stream by using the encoder's multi-channel multiplex and output functionality with cascaded transport stream input/output connections, as illustrated in Fig. 3.

4. Implementation

A photograph of the overall 4K/120 fps encoder with four 2K/120 fps encoder devices is shown in **Fig. 4**, and the encoder specifications are listed in **Table 1**. This encoder inputs a 4K/120 fps video with eight 3G-SDIs (third-generation serial digital interfaces), and outputs an HEVC temporal scalable stream in the MPEG-2 TS format by using a DVB-ASI

*2 MPEG-2 Transport Stream: A Moving Picture Experts Group standard digital container format for transmission and storage of audio, video, and Program and System Information Protocol (PSIP) data.

Fig. 4. Photograph of encoder.

Table 1. System specifications.

Video format	3840 × 2160 pixels / 120 fps
Video coding	HEVC Main 10 profile
Container format	MPEG-2 Transport Stream
Input	3G-SDI × 8
Output	DVB-ASI or 1000BASE-T

Fig. 5. Real-time encoding demonstration.

(digital video broadcasting - asynchronous serial interface) or by using Internet protocol (IP) connections. We observed that the 4K/120 fps bit streams encoded by our encoder were successfully decoded and played by other HFR systems.

A photograph taken during a demonstration of 4K/120 fps real-time transmission with our HFR encoder is shown in **Fig. 5**. Uncompressed 4K/120 fps images were encoded at a constant bit rate of 80 Mbit/s, where the bit rate of the base layer was 60 Mbit/s and that of the enhancement layer was 20 Mbit/s and images were distributed over an IP connection. The other real-time 4K/120 fps software decoder received the compressed bit stream, which was demultiplexed from MPEG-2 TS and decoded in real time. Then the 4K/120 fps decoded images were displayed on the screen by the 4K/120 fps-enabled projector.

No 120 fps video services have been available to date due to the enormous video data size and the requirement for compatibility with legacy systems. This has made it difficult to achieve 120 fps video distribution. Our device solves this problem and provides advanced high quality HFR video by combining it with other existing 4K/120 fps-enabled devices such as a camera, decoder, and projector.

5. Conclusion

We presented a new 120 fps real-time HEVC encoder for higher frame rate video encoding and transmission exploiting the existing HEVC encoder LSI. This 120 fps encoder makes it possible to achieve temporal scalable HEVC coding and transmission with upward compatibility for 60 fps. We plan to continue developing video coding devices to contribute to the provision of new services.

References

[1] M. Emoto, Y. Kusakabe, and M. Sugawara, "High-frame-rate Motion Picture Quality and Its Independence of Viewing Distance," J. Disp. Technol., Vol. 10, No. 8, pp. 635–641, 2014.

[2] T. Onishi, T. Sano, Y. Nishida, K. Yokohari, J. Su, K. Nakamura, K. Nitta, K. Kawashima, J. Okamoto, N. Ono, R. Kusaba, A. Sagata, H. Iwasaki, M. Ikeda, and A. Shimizu, "Single-chip 4K 60 fps 4:2:2 HEVC Video Encoder LSI with 8K Scalability," Proc. of 2015 Symposium on VLSI Circuits (VLSI Circuits), C54-C55, Kyoto, Japan, June 2015.

[3] ARIB STD-B32: "Video Coding, Audio Coding, and Multiplexing Specifications for Digital Broadcasting," Version 3.8, Sept. 2016.

Yuya Omori
Researcher, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in information and communication engineering from the University of Tokyo in 2012 and 2014. He joined NTT Media Intelligence Laboratories in 2014 and has been engaged in research and development (R&D) of parallel processing architecture and a high efficiency video codec algorithm. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Institute of Electronics, Information and Communication Engineers (IEICE).

Takayuki Onishi
Senior Research Engineer, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in information and communication engineering from the University of Tokyo in 1997 and 1999. He joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 1999 and has been engaged in R&D of high quality image coding and transmission. He is a member of IEEE and IEICE.

Hiroe Iwasaki
Senior Research Engineer, Supervisor, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

She received a B.S. and Ph.D. in information science from Tsukuba University, Ibaraki, in 1992 and 2006. She joined NTT Switching System Laboratories in 1992. She has been involved in R&D of an architecture and real-time operating system for multimedia embedded system LSIs. She moved to NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 1997. She is a member of IEEE, IEICE, and the Information Processing Society of Japan.

Atsushi Shimizu
Senior Research Engineer, Supervisor, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in electronic engineering from Nihon University, Tokyo, in 1990 and 1992. Since joining NTT in 1992, he has been working on video compression algorithm and software development. He is a member of IEICE, the Institute of Image Electronics Engineers of Japan, and the Institute of Image Information and Television Engineers.

Surround Video Stitching and Synchronous Transmission Technology for Immersive Live Broadcasting of Entire Sports Venues

Takako Sato, Koji Namba, Masato Ono, Yumi Kikuchi, Tetsuya Yamaguchi, and Akira Ono

Abstract

NTT Service Evolution Laboratories is promoting the research and development of the live immersive reproduction of entire spaces and environments by simultaneously transmitting video, audio, and related information from a remote location. In this article, we propose a technique to combine multiple 4K video pictures horizontally and vertically in real time to create a panorama video with high resolution and a wide viewing angle (at least 180°)—much higher than the resolution that can be achieved with a single camera. We also introduce technology to synchronously transmit other video and audio data to remote venues with low latency in addition to the stitched video.

Keywords: video stitching, media synchronous transmission, low latency

1. Introduction

NTT Service Evolution Laboratories is actively driving research into the live immersive reproduction of entire spaces and environments by simultaneously transmitting live video, audio, and related information from the source location. By establishing technologies for transmitting immersive live video to remote locations, we aim to provide immersive public viewing services that can provide captivating experiences of events such as sports matches and live shows; the goal is to enable viewers to experience such events as though they were actually at the venue.

In this article, we introduce surround video stitching/synchronous transmission technologies that can transmit live video of an entire sports event taking place at, for example, a large sports field, and reproduce it at a remote location so that it appears as if the action was taking place right in front of the audience. This technology can also be applied to public viewing for productions such as theater shows and musical concerts that take place on a large stage.

2. Surround video stitching/synchronous transmission technologies

In recent years, Hi-Vision and 4K video have made inroads in a wide variety of fields, and they are being increasingly incorporated into the development of higher-resolution 8K services, mainly in broadcasting applications. However, these services are not yet able to provide complete, unbounded views of sporting events taking place on large sports fields (e.g., baseball, soccer, and rugby). We are working to establish surround video stitching and synchronous transmission technologies that offer the real-time stitching and transmission of high-resolution video (beyond 4K/8K) covering a wide viewing angle so that people can view events as if they were actually there. These technologies consist of two core

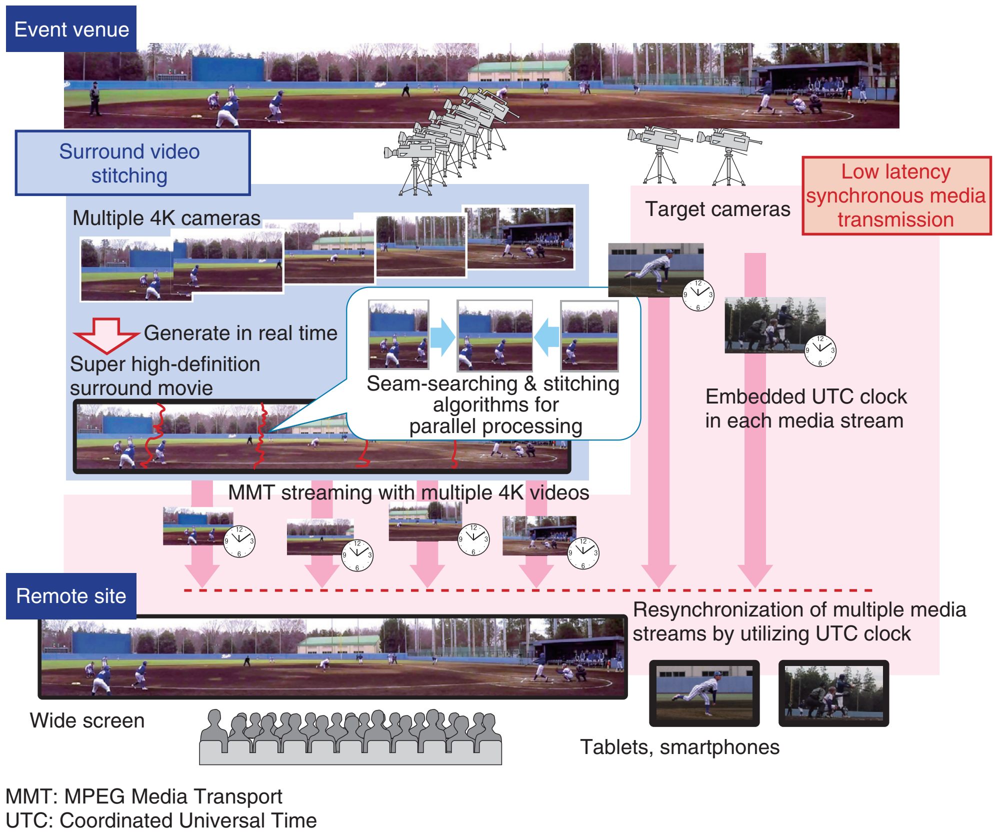


Fig. 1. Surround video stitching/synchronous transmission technologies.

components (**Fig. 1**):

- Surround video stitching: This combines multiple 4K video images both horizontally and vertically in real time to produce high-definition surround video with a wide viewing angle (180° or more) that greatly exceeds the resolution of images that can be taken with a single camera.
- Low latency synchronous media transmission: Surround video covering the entire sporting venue is partitioned into multiple images and combined with sounds from the event and close-up videos of the players, and then synchronously transmitted and played back with low latency on large-scale Hi-Vision systems at remote venues or on spectator tablet devices. MMT (MPEG* Media Transport) is used to transmit the content, and performing delay compensation according to the playback environment makes it possible to accurately synchronize multiple video and audio sources even on disparate playback devices.

3. Surround video stitching technology

In this section, we explain the various elements of surround video stitching technology.

3.1 Real-time video stitching issues and solutions

Various conventional algorithms can be used to create 4K video panoramas and 360° views, but these are all centered on techniques that demand off-line processing. For real-time processing of multiple 4K video images, amounting to approximately 15 gigabits of data per second, it is necessary to overcome many technical issues in terms of improving the processing speed. When creating panoramas from multiple video images, correction processing (homography transformation) is performed on each video image to reduce the effects of parallax arising from the use of multiple cameras. The stitching lines

* MPEG: Moving Picture Experts Group. A working group established by ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission) to develop standards for digital audio and video compression.

(seams) between adjacent video images are then located by considering the movements and other attributes of people and other moving objects in the video, and finally, the multiple video images are merged to yield panorama videos. The seam positions are determined so as to avoid moving objects and thus prevent artifacts such as the stretch and truncation of moving objects that occur when moving objects cross a seam.

Our surround video stitching technology makes it possible to stitch, in real time, multiple 4K/60 fps video images that adjoin one another in the horizontal and vertical directions. The computationally intensive process of seam search, which involves video analysis, is conducted on multiple servers in parallel. By combining eleven 4K images on the horizontal plane, we can generate surround video with a 180° viewing angle. We describe below the real-time processing operations in more detail and introduce a camera rig that can shoot video appropriate for stitching.

3.2 Faster seam search

In searching for seams, information such as luminance differences (motion information) and edge intensities (shape information) obtained by analyzing the video frames is used, but in our surround video stitching technology, this analysis has to be performed in parallel in order to achieve sufficient speed. It is possible to use information from previous video frames to stabilize seam positioning so that the seams do not jump about unnecessarily, even for fast-moving objects. However, when video analysis and seam search are performed in parallel, it may not be possible to refer to the seam information of previous video frames. To address this problem, we perform seam detection processing on scaled-down video images to obtain approximate seam information for previous video frames, thereby achieving both high processing speeds and accurate seam detection.

To combine multiple 4K images that are tiled both horizontally and vertically, we have to perform stitching processing in many overlapping regions (**Fig. 2**). Since the sequential execution of stitching processing in the horizontal and vertical directions would increase the overall processing time, the seam search processing in each overlapping region is performed in parallel. If we search for seams separately in each area, it is possible that these seams will not line up properly from one region to the next. This makes it necessary to share information about the seam endpoints between adjacent areas so that continuous seams can be formed between regions in the horizontal and vertical directions (**Fig. 3**).

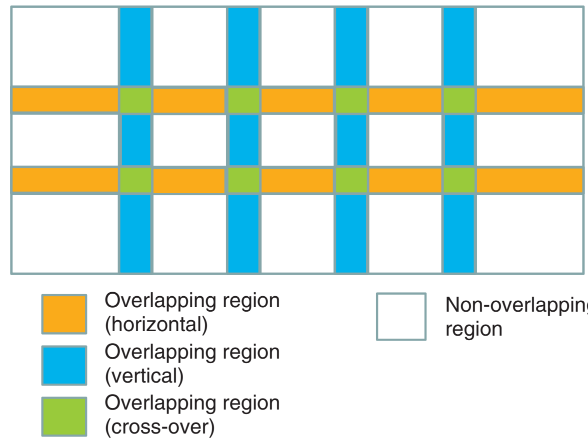


Fig. 2. Example of the overlapping regions used in 5(H) × 3(V) 4K video stitching.

3.3 Synchronous distributed processing architecture

With our surround video stitching technology, the abovementioned correction processing and seam search/stitching processing are done across multiple servers in parallel, enabling even huge video images to be processed in real time. To finally merge the video images processed by each server into a single 4K × N surround video, we need to synchronize the various processes performed on each video frame and output the processing results synchronously. In this technology, synchronous control is facilitated by applying the same time stamp to video frames captured at the same time before they are distributed among the servers.

Instead of generating new time stamps every time a video frame is input, it is possible to generate time stamps from the time codes that are added by cameras or superimposed on the input video. Conversely, when the stitched video is output, the original time codes can be recovered and superimposed on the picture. In this way, we can propagate time codes assigned by cameras or other equipment directly to the later stages of our system (e.g., encoding or transmission equipment), enabling stitched video to be played back in perfect synchronization with other video materials that are transmitted directly without surround video stitch processing.

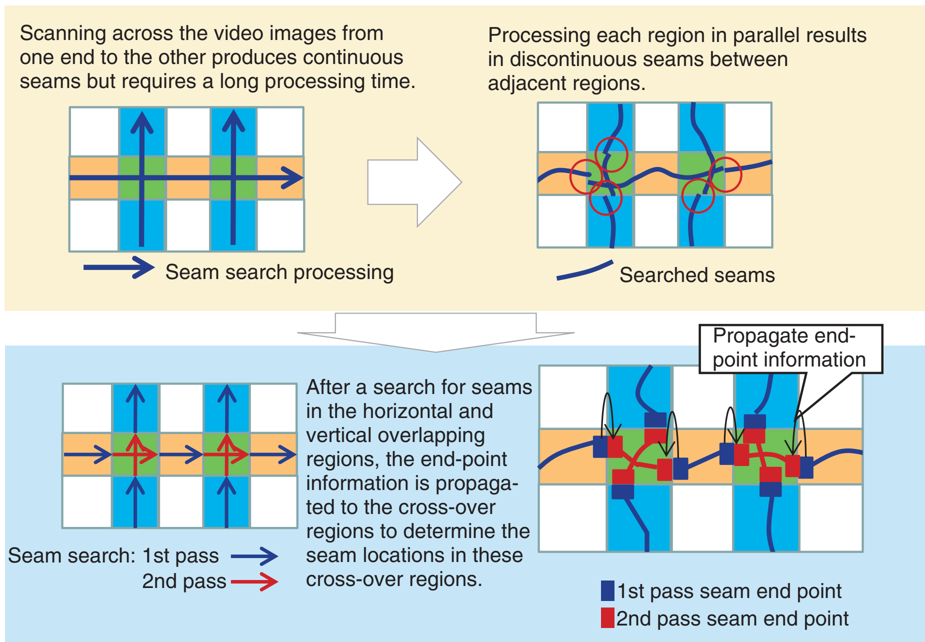


Fig. 3. Parallel processing to ensure seam continuity.

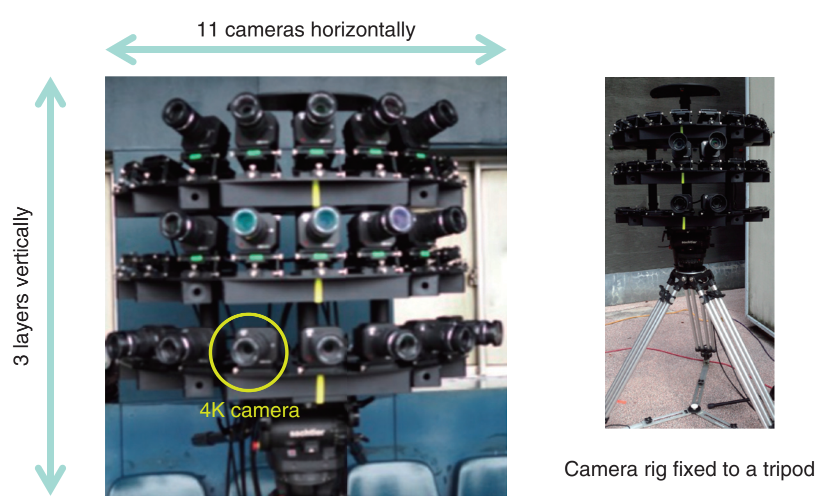


Fig. 4. Video camera rig with multiple 4K cameras mounted horizontally and vertically.

3.4 Precise video capture from multiple 4K devices

Before the video stitching is done, the effects of parallax arising from the use of multiple cameras are reduced by performing correction processes such as homography transformation. However, it is not possible to completely eliminate the effects of parallax in all regions of the videos because they include depth and movement. To create stitched images of the highest quality, we must accurately adjust the camera positions so as to minimize parallax in the source material. We have developed a video capture rig that can quickly and accurately adjust camera positions in the horizontal and vertical directions, and we use this rig to reduce the video capture workload and improve the quality of the stitched video (**Fig. 4**).

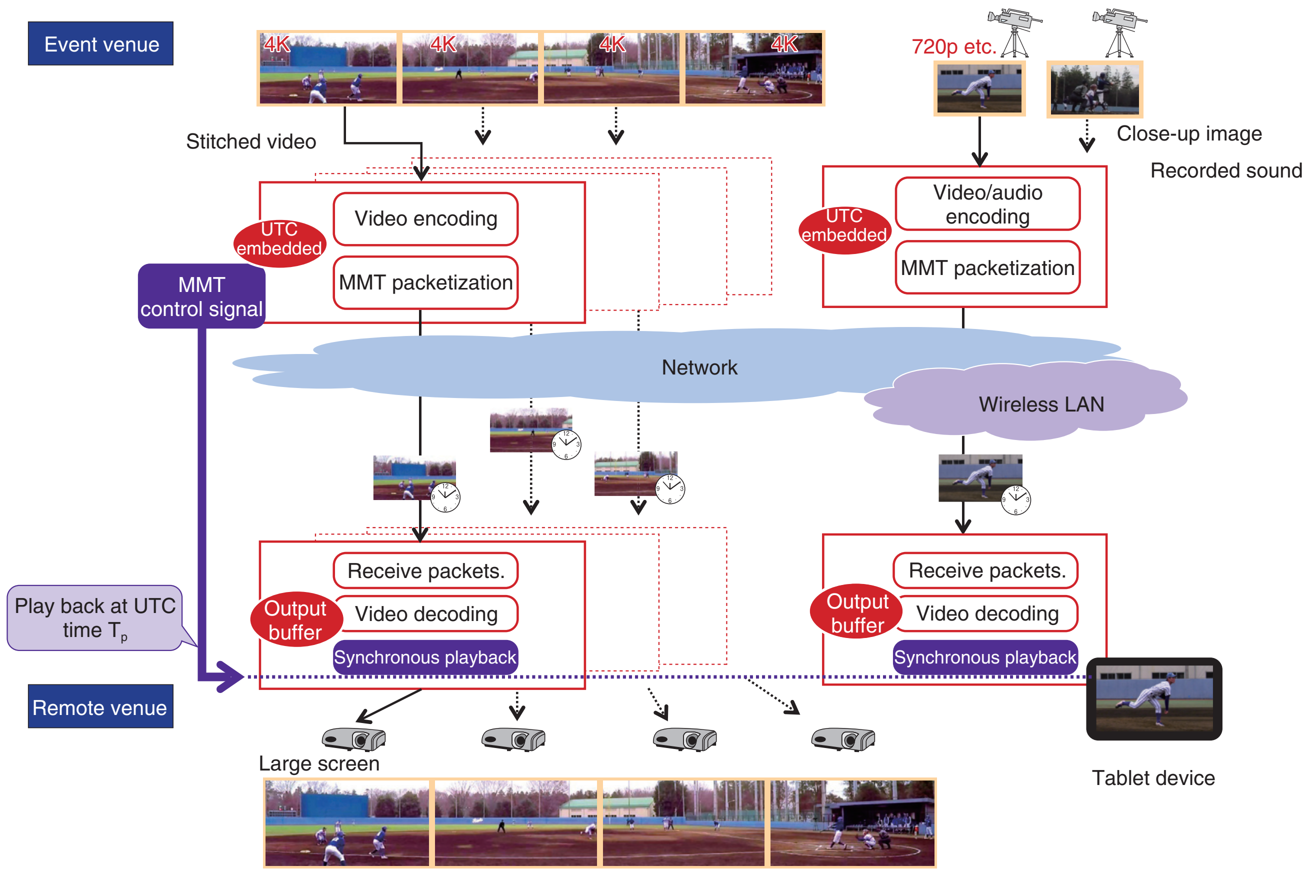


Fig. 5. Low-latency media synchronization and transmission.

4. Low-latency media synchronization and transmission technology

A variety of different display devices may be used when viewing content at a remote location, including large-scale displays and personal tablet devices. Content may also be delivered via a transmission network that includes various paths such as leased lines, the Internet, and wireless local area networks (LANs). If multiple video and audio signals are transmitted across these mixed environments, the playback timing of individual video and audio streams will vary due to differences in the propagation delay and processing latency on each route.

Our low-latency media synchronization and transmission technology achieves low-latency transmission while controlling the playback terminal buffer and using MMT-conformant synchronization control signals based on UTC (Coordinated Universal Time) to suppress the variations in the multiple video and audio sources. MMT also specifies the use of error correction coding to accommodate the loss of packets in transit. We implement error correction using a lightweight low-latency coding technique called LDGM (low density generator matrix) [1] developed by NTT Network Innovation Laboratories to enable stable synchronous playback even when close-up videos are transmitted to tablet devices via a wireless LAN.

The low-latency media synchronization processing flow is described below, taking the transmission of surround video as an example (**Fig. 5**). The surround video is transmitted after it has been partitioned into 4K units and encoded using a technique such as HEVC (High Efficiency Video Coding). During this process, the playback time, T_p, is transmitted to the display terminal as a UTC time stamp together with the video signal. Once the display terminal has received and decoded the video, it is stored in a buffer region until time T_p and then synchronously played back. Our low-latency media synchronization and transmission technology allows the buffer content to

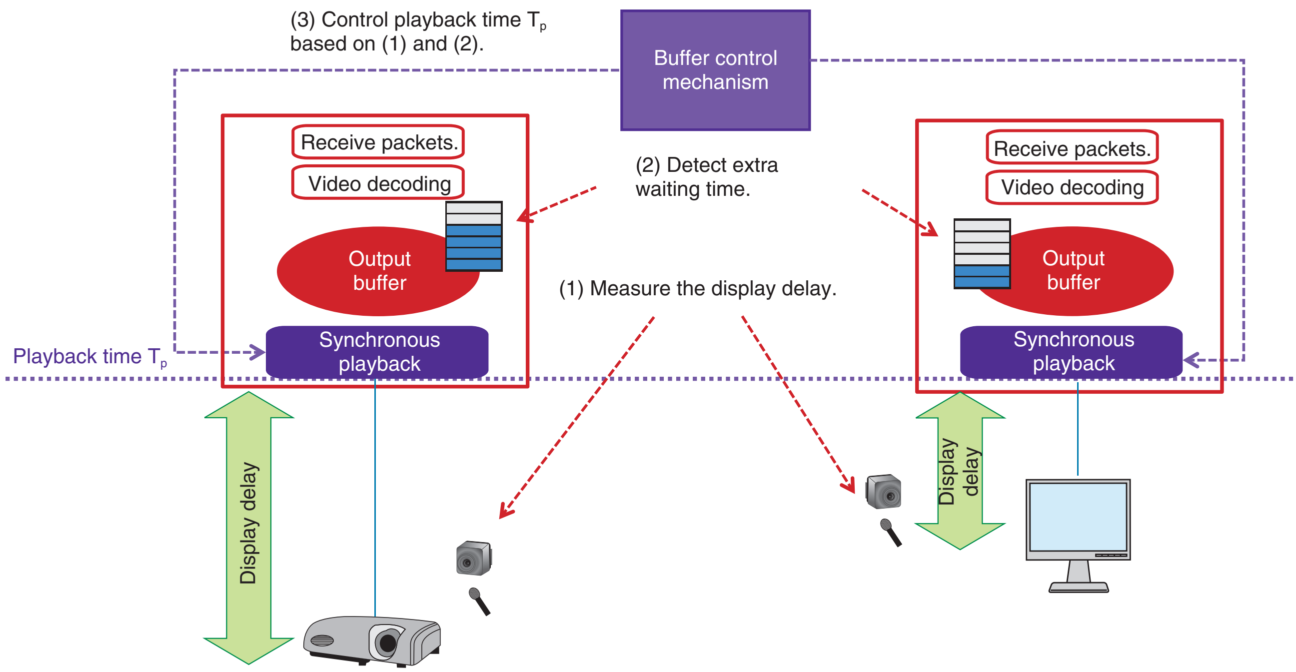


Fig. 6. Controlling the playback buffer region.

be controlled according to the playback environment so as to simultaneously achieve 1) guaranteed synchronization in different playback environments and 2) minimal video transmission delays (**Fig. 6**).

4.1 Guaranteed synchronization in different playback environments

In media transmission using MMT, synchronization of the playback timing can be guaranteed at the stage where video is output from the decoder, but the synchronization can still deviate due to differences in playback environments. This is because the time at which the video is actually presented is later than T_p due to differences in the display delays of different playback devices (projectors, monitors, tablets, etc.) and the delays of processes such as geometry correction in projectors, which are liable to vary. This can be fixed by measuring the delay in each playback environment and controlling a buffer to increase or decrease the delay so that the playback timing is correctly synchronized to T_p in each playback environment, thereby facilitating synchronized playback even in different playback environments.

4.2 Reduced latency of video transmission

The buffer that holds the decoded video until time T_p is needed for synchronization of video playback timings, but when this buffer contains more space than necessary, the video is delayed longer than necessary, resulting in additional latency. To avoid this, the amount of buffered video can be measured during playback, and by adjusting the buffer to eliminate excess capacity, we can achieve low-latency playback.

5. Future prospects

With the aim of implementing truly immersive public viewing services, NTT Service Evolution Laboratories is conducting research and development aimed at enhancing the real-time high-resolution wide-angle video stitching technologies we have developed so far and implementing functions with lighter system overheads and enhanced video processing to expand the application fields.

Reference

[1] T. Nakachi, T. Yamaguchi, Y. Tonomura, and T. Fujii, “Next-generation Media Transport MMT for 4K/8K Video Transmission,” NTT Technical Review, Vol. 12, No. 5, 2014.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201405fa6.html

Takako Sato
Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

She received a B.S. in mathematics from Hirosaki University, Aomori, in 1994. She joined NTT Multimedia Systems Department in 1994 and developed video transmission systems. She was also involved in the development of IP-based broadcasting systems. She moved to NTT Service Evolution Laboratories in 2012 and is developing new services based on R&D products.

Koji Namba
Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received an M.E. in electronic engineering from Osaka University in 1999. He joined NTT in 1999 and developed digital rights management systems. From 2002 to 2015, he worked at NTT WEST, where he was involved in network engineering projects and the development of consumer services. He moved to NTT Service Evolution Laboratories in 2015, where he has been developing immersive telepresence technology.

Masato Ono
Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received an M.E. in information engineering from University of Tsukuba, Ibaraki, in 2006. He joined NTT EAST in 2006 and engaged in network engineering, customer service management, and creating business-to-business services. He moved to NTT Service Evolution Laboratories in 2016 and is developing new services based on immersive telepresence technology.

Yumi Kikuchi
Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

She received a B.E. and M.E. in electrical engineering from Tokyo University of Science in 2000 and 2002. Since joining NTT in 2002, she has been researching and developing content configuration methods and content delivery systems. Her present research is focused on feature point extraction of images and matching using the extracted feature points.

Tetsuya Yamaguchi
Senior Research Engineer, Supervisor, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E., M.E., and Ph.D. in information engineering from Osaka University in 1997, 1999, and 2008. Since joining NTT in 1999, he has been researching and developing content distribution and navigation systems. His current interests are advanced media transport and ultra-realistic communications.

Akira Ono
Senior Research Engineer, Supervisor, Natural Communication Project, NTT Service Evolution Laboratories.

He received an M.E. in computer engineering from Waseda University, Tokyo, in 1992. He joined NTT in 1992 and engaged in research and development of video communication systems. From 1999 to 2010, he worked at NTT Communications, where he was involved in network engineering and creating consumer services. He moved to NTT Cyber Solution Laboratories (now, NTT Service Evolution Laboratories) in 2010. Since 2015, he has been studying the immersive telepresence technology called "Kirari!".

Powerful Sound Effects at Audience Seats by Wave Field Synthesis

Kimitaka Tsutsumi and Hideaki Takada

Abstract

Sound spatialization technology has become increasingly common within movie theaters and other venues for creating sound effects coming from all directions. NTT Service Evolution Laboratories has driven research into sound spatialization technology using a linear loudspeaker array to enhance the sense of reality in Kirari!, NTT's immersive telepresence technology. We have developed a system that enables the creation of sound images in front of loudspeakers that could not be created using conventional surround-sound systems. This article introduces the theory and implementation of wave field synthesis technology using a linear array of speakers for generating sounds that close in on venue seats.

Keywords: wave field synthesis, sound spatialization, spatial acoustics

1. Introduction

NTT Service Evolution Laboratories has been conducting research and development (R&D) of an immersive telepresence technology named "Kirari!" to replicate the reality of a sporting venue in real time at remote venues located anywhere in the world. Kirari! is a technology for showing a sports match as if it were happening in front of viewers by synchronously presenting virtual images and sound images of players consistent with their locations on a holographic display. In the case of soccer and judo, synchronous creation of sound images such as a kicked soccer ball and thrown judo players consistent with the locations of projected images provide higher reality to an extent that has been impossible using conventional stereo systems. For this reason, sound localization technology that creates virtual sound sources at any location within a venue is finding widespread use in movies and sports matches requiring a high sense of presence or powerful scenes. It is becoming essential for drawing out a highly realistic sensation using acoustics.

At NTT R&D Forum 2016, we used virtual speakers for sound localization by radiating ultrasonic waves in a controlled fashion at subjects on a screen. Since the audience would hear that sound reflected off the screen, they would feel the sound as if it were coming directly from the subjects on the screen. However, control of the sound image was limited to left/right positions, which meant the results were hardly different from the sound produced by stereo reproduction. Another problem was that the use of reflected ultrasonic waves prevented the output of high sound pressure levels.

In response to these problems, we have investigated wave field synthesis technology capable of achieving sound-image control in the depth direction and consequently achieved sound reproduction with a high sense of presence in which sound can close in on venue seats. In addition, wave field synthesis can be implemented with standard speakers, so a large volume of sound can be produced relative to ultrasonic waves. This makes it possible to demonstrate a high-presence effect even in large venues. Wave field synthesis technology has been incorporated in Kirari! and successfully applied in the field at many events including NTT WEST Group Collection 2016, NTT R&D Forum 2017, Kabuki Virtual Theater, and Niconico Cho Kabuki 2017.

In this article, we present an overview of wave field synthesis technology and introduce the system used in the exhibition of this technology at Niconico Cho Kabuki 2017 as a field trial.

Fig. 1. Loudspeaker array.

2. Wave field synthesis technology

In this section, we explain the background of this technology and report on its application. We also describe problems that need to be addressed.

2.1 Theory of wave field synthesis

Wave field synthesis is sound-reproduction technology that reproduces a sound's wave front based on a physical model. It is achieved according to theory based on the Kirchhoff-Helmholtz integral equation [1]. This equation means that any sound field can be reproduced by controlling the sound pressure and sound-pressure gradient at the boundary surface enclosing that space. In essence, this means that if the walls, ceiling, and floor can be filled with infinitesimal speakers and if those speakers can be controlled, any sound field can be reproduced. However, if this were attempted using the Kirchhoff-Helmholtz integral equation in its original form and using commercially available speakers, it would require a large number of speakers with special characteristics that are difficult to achieve, and the total number of speakers needed would be massive.

We therefore decided on an implementation using an approximation based on Rayleigh's first integral and limiting the target sound field to a plane. In actuality, we form a speaker array with many closely spaced, linearly arranged speakers, set the target sound field, and synthesize it on the same plane as the speaker array (**Fig. 1**).

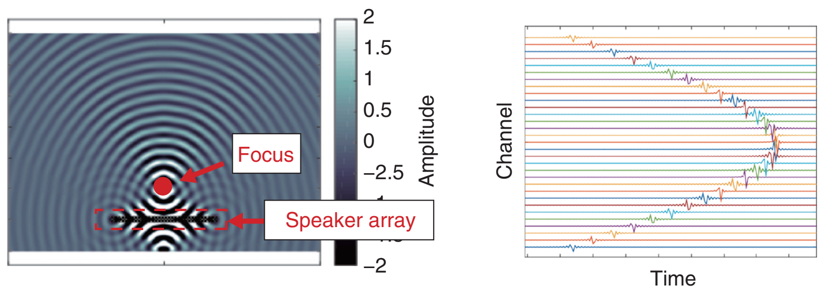
2.2 Kirari! and wave field synthesis technology

The wave field synthesis technology studied at the NTT laboratories up to now should actually be called wave field reconstruction technology; this technology uses a speaker array to reproduce with high fidelity a sound field recorded with a microphone array [2]. Here, the sound field in the recording venue can be reproduced in its original form in the performance venue by processing and regenerating the recorded sound without having to detect the number and location of sound sources. However, because it is necessary to arrange the microphone array near the sound sources in the recording venue, the array may then appear in the video that is being synchronously recorded, making this setup difficult to use in an actual live broadcast. In addition, the difficulty here of separating individual sound sources detracts from the flexibility of limiting sounds to only specific ones and adding editing.

Considering, therefore, that direction and editing may be applied to content, we studied and adopted a new type of wave field synthesis technology that produces virtual sound sources from monaural sound sources. Editing and producing content after recording individual sound sources using a directional gun microphone as used in ordinary recordings and then adding sound spatialization effects to the monaural sound sources constitute a technique tailored to a realistic workflow.

2.3 Focused source method

Setting a target sound field so that the sound emitted from speakers creates a focus in front of the speaker array makes it possible to achieve a virtual sound source that pops up at this focus that is positioned nearer to the audience seats than the installed



(a) Positional relationship between speaker and focus

(b) Output audio for each speaker when inputting a pulse

Fig. 2. Sound field creating a focused sound source and audio input to each speaker.

speakers. This sound-reproduction technique is called the focused source method [3].

When the focused source method is used, sound power can be concentrated at the point of focus by exploiting a digital filter determined analytically from the positions of the focus and speakers. Since the filters differ for each speaker, a monaural sound source is sufficient to obtain multichannel output even if using a large number of speakers. The role of the digital filter is to adjust the timing of sound reproduction and power at each speaker so that the sounds from every speaker arrive exactly at the same time at the focus position.

The positional relationship between the speaker array and the point of focus is shown in **Fig. 2(a)**. The audio output from each speaker in the case of a pulse input signal is shown in **Fig. 2(b)**. In this example, 32 speakers are lined up at 0.1-m intervals, and a focus is created 2 m from the center of the speaker array. Since the distance from either end of the speaker array to the focus is longer than the distance from the center of the speaker array to the focus, generating sound from these speakers simultaneously would cause the sound emitted from the center of the speaker array to pass the focus before the sounds emitted from both ends of the speaker array arrive at the focus. Consequently, to concentrate sound power at the focus, sound reproduction should begin at the speakers furthest away from the focus. On examining the output from each speaker shown in Fig. 2(b), it can be seen that pulse generation begins at the speakers at each end of the array that lie furthest from the focus.

Next, we show the reproduced wave front when the sounds emitted from each speaker come together. If the sound source is placed at the focus position and a single pulse is generated, the wave front will spread out in the shape of concentric circles (**Fig. 3**). The reconstruction of the desired output from the speaker array to generate a focus at the same position as above is shown in **Fig. 4**. It can be seen here that wave fronts similar to that shown in Fig. 3 are formed where the sounds emitted from multiple speakers overlap. Using a digital filter determined by the focused source method as described above enables the generation of virtual sound sources in any location, including the area on the audience-seating side of the speaker array.

2.4 Problems in implementation

Unfortunately, using wave field synthesis does not enable sound effects that close in on the audience from the screen to be experienced from every seat in the venue. The theory of wave field synthesis assumes the use of an infinitely long speaker array, but this is not possible in practice since speaker arrays are actually formed by lining up a finite number of commercially available speakers. Furthermore, at an actual event, regulations such as fire laws or other constraints dictate where speakers can be installed, so it is not unusual to have to open up a gap for a passageway in a speaker array. This causes narrowing of the sweet-spot area in which the position at which an audience member perceives the sound image matches the position of a focused sound source set beforehand. As a result, the number of seats from which audience members can simultaneously experience the effects of pop-up sounds becomes smaller.

In response to this problem, we introduced a technique that creates a bend at a point in the linear speaker array to form a segmented array in which each segment acts as an independent array and that

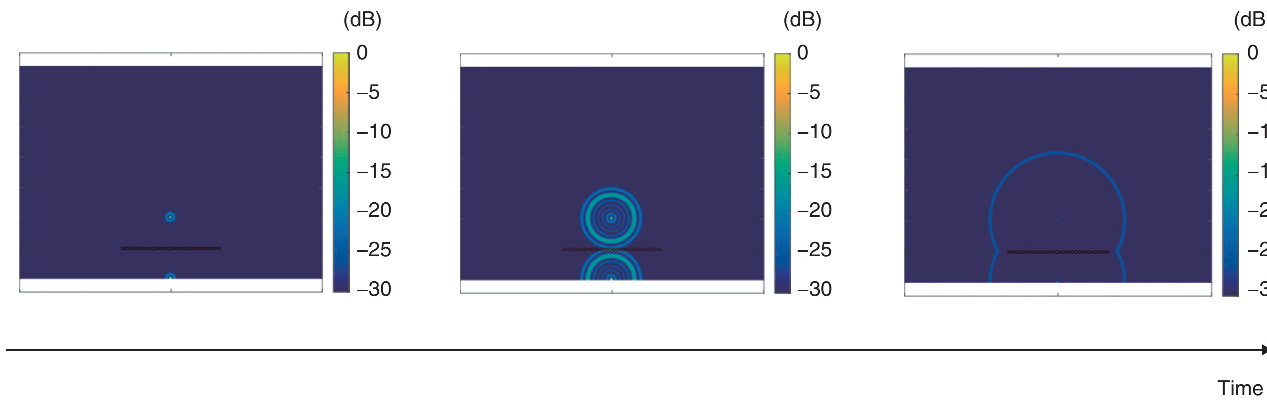


Fig. 3. Wave front of monopole sound source.

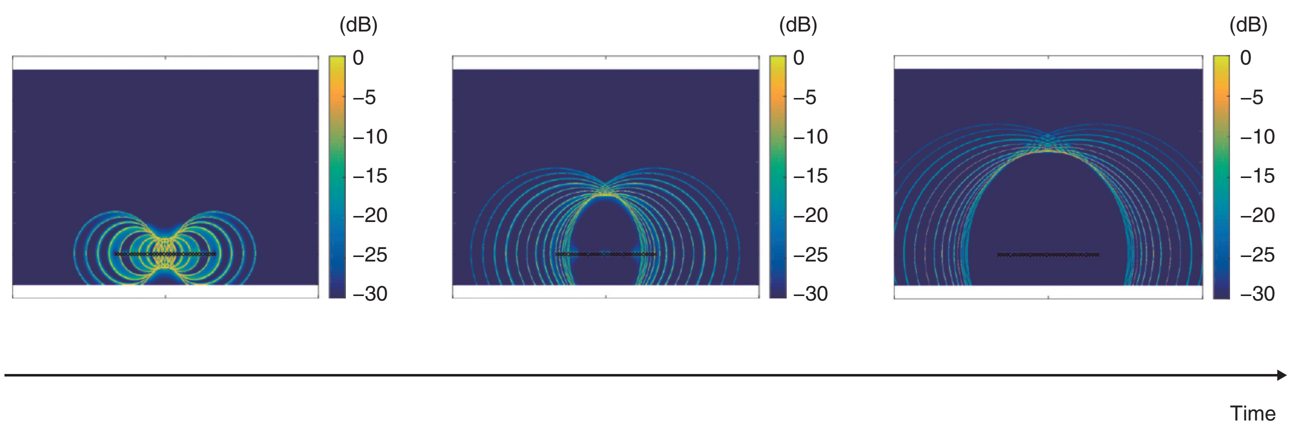


Fig. 4. Reproduced wave front using wave field synthesis.

generates a focused sound source for each segment, as shown in **Fig. 5** [4]. In this way, we succeeded in expanding the sweet-spot listening area. At present, research on this sweet-spot problem is progressing around the world, and we as well are working to further expand the sweet-spot listening area by pursuing theoretical advances in wave field synthesis.

3. Field trial: Niconico Cho Kabuki 2017

We introduced special effects at the Niconico Cho Kabuki 2017 event using 240 channels and speaker arrays with a total length of 36 m to perform wave field synthesis targeting 350 seats on the first level of the venue (**Fig. 6**). Specifically, we prepared sound effects achieved through wave field synthesis in a sound-reproduction personal computer (workstation) beforehand and had an operator activate these effects in unison with the actor's performance as a system for reproducing sound effects.

The overall system configuration is shown in **Fig. 7**. This system sends out audio data from two audio interfaces connected in parallel at the workstation and reproduces audio from speakers with built-in amplifiers so that sound effects achieved by wave field synthesis resound in the venue. In this way, sound achieved through wave field synthesis was used to create effects for various scenes in the kabuki performance. For example, in the scene where a stone lantern explodes and a dragon flies away, we used the system to reproduce the sound of stone lantern fragments scattered by the explosion falling onto the audience at their seats, and in the final scene, we used it to reproduce the sound of a burning house collapsing near the audience. Generating the sound of falling debris or stones near the ears of audience members in

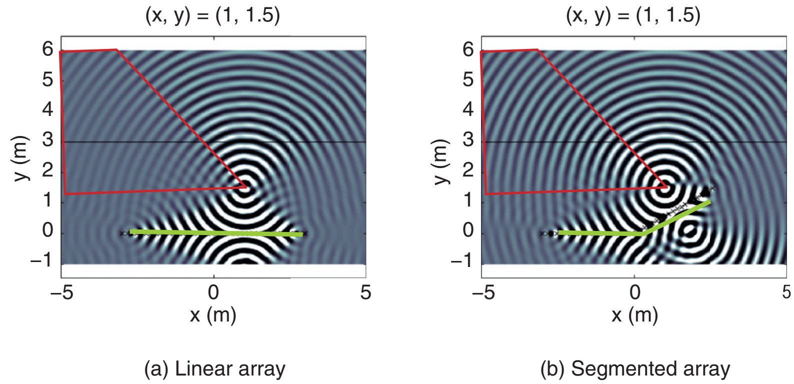


Fig. 5. Extended sweet spot. The area within the red border in (b) shows that the radiated sound from the focused source is stronger than the corresponding area in (a). This means that a larger area of the desired sound field is reproduced by a segmented array than by a linear array.

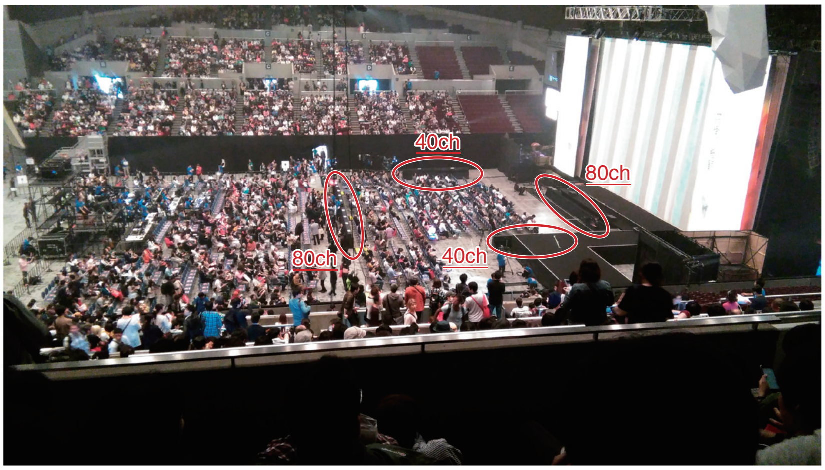


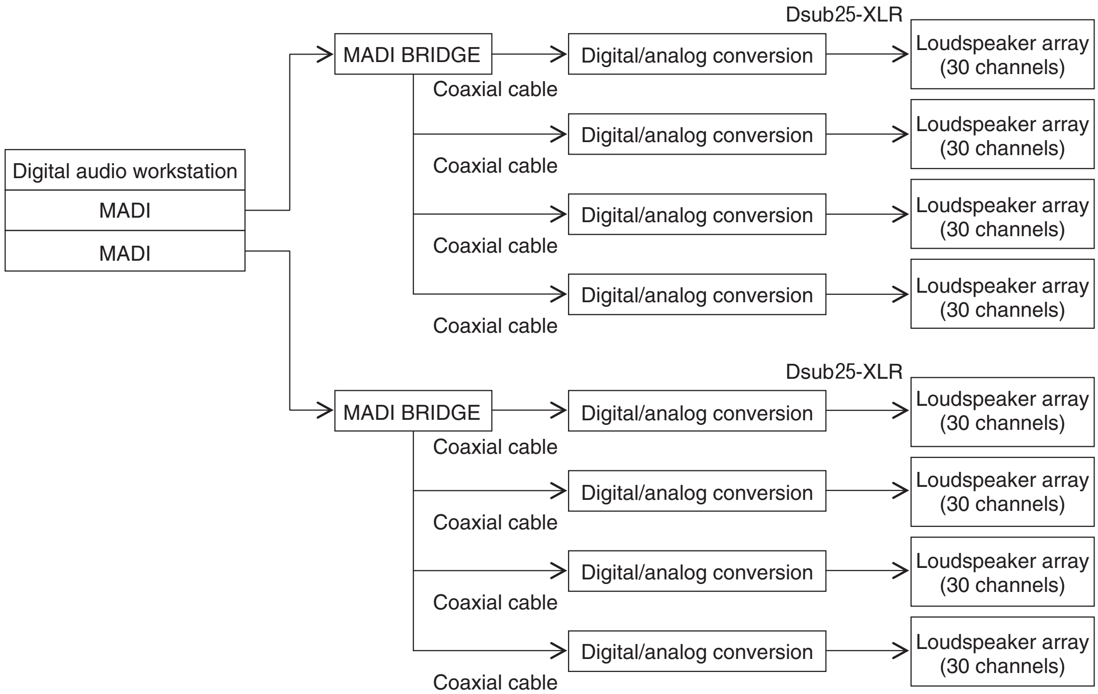
Fig. 6. Field trial at Niconico Cho Kabuki 2017.

this way achieved a high sense of presence as if debris was actually falling inside the venue.

Despite the fact that more than 5000 people attended each performance at this venue, Makuhari Messe Event Hall, constraints in installing equipment meant that only a portion of the audience seated on the first level were able to experience the special sound effects. Nevertheless, we succeeded in introducing special effects using wave field synthesis at a live event held in a large hall, and we received positive feedback on the effective and rich expression that we achieved in combination with the performance while dealing with the sweet-spot problem. This trial provided our team with an extremely valuable experience.

4. Future development

The use of a large number of speakers in systems applying wave field synthesis can create virtual sound images between loudspeakers and the audience not possible with surround-sound systems. However, there are still many problems to be addressed even with the use of a loudspeaker array, including the sweet-spot problem in the focused source method. It is therefore essential to continue researching wave field synthesis. Additionally, to provide home users with diverse content exploiting sound spatialization effects, we need to investigate sound spatialization technology using headphones or a small speaker array that can achieve the same effects as wave field



MADI: multichannel audio digital interface

Fig. 7. System configuration.

synthesis using a loudspeaker array. Here, research results in wave field synthesis are useful as benchmarks for such studies. By working on these problems in parallel, we will be able to research sound-reproduction technologies that achieve high reality for both theater and home use.

References

[1] K. Furuya and S. Koyama, "Wave Field Synthesis; Principles and Applications," Journal of the Institute of Image Information and Television Engineers, Vol. 68, No. 8, pp. 621–624, 2014 (in Japanese).

[2] S. Koyama, K. Furuya, H. Uematsu, Y. Hiwasaki, and Y. Haneda, "Real-time Sound Field Transmission System by Using Wave Field Reconstruction Filter and Its Evaluation," IEICE Trans. Fundamentals., Vol. E97-A, No. 9, pp. 1840–1848, 2014.

[3] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, "Physical and Perceptual Properties of Focused Sources in Wave Field Synthesis," Proc. of 127th AES Convention, paper 7914, New York, USA, Oct. 2009.

[4] K. Tsutsumi and H. Takada, "A Method Using Auxiliary Arrays for Extending Listening Area of Focused Source in Wave Field Synthesis," Proc. of the 2017 Spring Meeting of the Acoustical Society of Japan, pp. 433–436, Kawasaki, Kanagawa, Japan, Mar. 2017 (in Japanese).

Kimitaka Tsutsumi
Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E. and M.E. in informatics from Kyoto University in 2004 and 2006. He joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 2006. He has participated in standardization efforts in International Telecommunication Union Telecommunication Standardization Sector (ITU-T). He joined NTT DOCOMO Research Laboratories in 2010, where he was involved in standardization activities of 3GPP (3rd Generation Partnership Project) on enhanced voice services. He has been with NTT Service Evolution Laboratories since 2015. His current research interests include speech coding and signal processing for spatial reproduction. He received the Technical Development Award from the Acoustical Society of Japan (ASJ) in 2016. He is a member of ASJ, the Audio Engineering Society, and the Institute of Electrical and Electronics Engineers (IEEE).

Hideaki Takada
Senior Research Engineer, Supervisor, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E. in information and communication technology from Tokai University, Shizuoka, in 1995, an M.E. in information systems from the University of Electro-Communications, Tokyo, in 1997, and a D.S. in global information and telecommunication studies from Waseda University, Tokyo, in 2007. He joined NTT Integrated Information & Energy Systems Laboratories in 1997. He has been researching three-dimensional (3D) visual perception and 3D display technology. He received the Best Paper Award from the 3D image conference in 2001 and 2005, the Young Researcher's Award from the Virtual Reality Society of Japan in 2002, the Achievement Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2003, a Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology of Japan in 2006, and Technical Committee Prize Paper Awards from IEEE in 2014 and 2016. He is a member of IEEE and the Institute of Image Information and Television Engineers.

Smooth Motion Parallax Glassless 3D Screen System that Uses Few Projectors

Motohiro Makiguchi and Hideaki Takada

Abstract

We propose a glassless 3D (three-dimensional) screen system that enables natural stereoscopic viewing with motion parallax in a wide viewing area. The use of a spatially imaged iris plane screen and linear blending technology makes it possible to use fewer image sources and projectors than existing multi-view projection systems.

Keywords: glassless 3D, optical linear blending, motion parallax

1. Introduction

A number of methods have been proposed for glassless three-dimensional (3D) displays that include motion parallax. One method, for example, involves projecting images from multiple directions onto a special screen having a narrow diffusion angle to create a multi-view image (**Fig. 1(a)**). Natural 3D vision can be achieved with this method with only a slight decrease in resolution [1, 2]. Jones et al. [2] proposed the use of this method to achieve a natural 3D vision system including motion parallax in a 135-degree viewing range by arranging 216 projectors at 0.625-degree intervals. This system can project images of people with high presence as if they were actually in that location.

However, in systems like this using many projectors, several things are necessary in order to switch viewpoints smoothly. These include creating many multi-view video sources, preparing the large number of projectors, and synchronizing the video sources. NTT Service Evolution Laboratories has been working to address these needs by focusing on optical linear blending technology that smoothly blends the luminance ratio of multiple images as the viewpoint moves (**Fig. 1(b)**). This is a glassless 3D screen system enabling smooth movement of viewpoints using very few projectors. In our earlier implementation [3], we found that it was difficult to increase the display area and widen the viewing area as a result of the optical configuration of the lens system.

In this article, we propose an improved multi-viewpoint glassless 3D display system that overcomes previous display and viewing area problems by using a special optical screen called a *spatially imaged iris plane screen* developed in collaboration with Tohoku University [4, 5].

2. Glassless 3D display screen adapted for viewpoint movement

In this section, we describe the optical linear blending technology and the spatially imaged iris plane screen.

2.1 Optical linear blending technology

Changing the luminance ratio of two overlapped objects separated by a fusion limit enables the position of an object perceived by a person to change smoothly between the two objects (**Fig. 2**). Utilizing this phenomenon makes it possible to present depth using motion parallax and binocular disparity. For example, let us assume a system such as that shown in **Fig. 3**, in which the luminance of Object 1 gradually decreases as the viewpoint moves from Viewpoint 1 to Viewpoint 2 and the luminance of Object 2

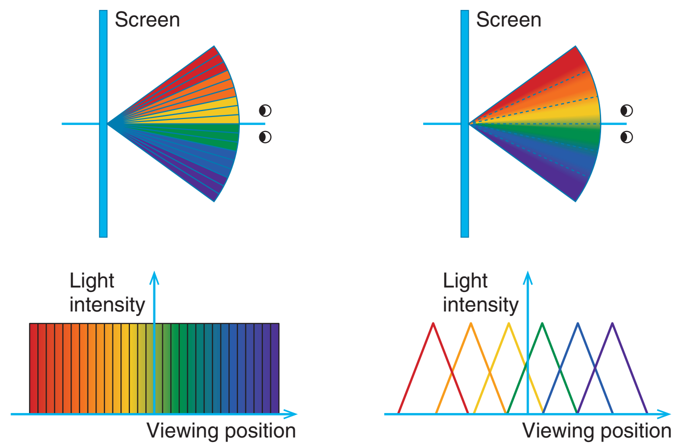


Fig. 1. (a) Conventional super multi-view 3D screen; (b) our target.

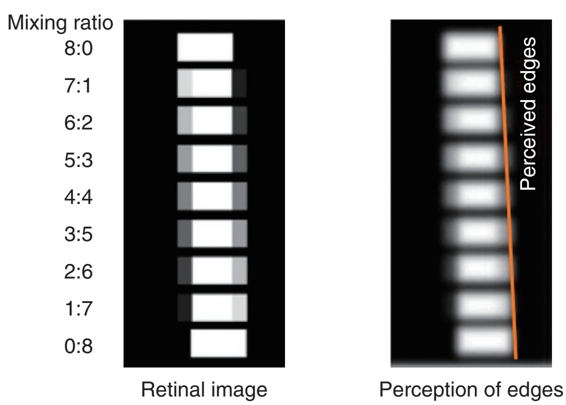


Fig. 2. Principal of linear blending.

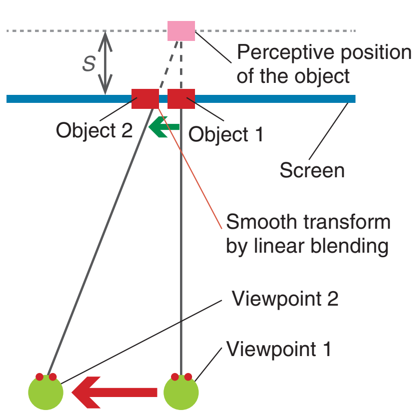


Fig. 3. Depth perception by motion parallax.

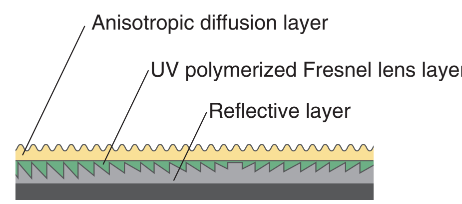


Fig. 4. Structure of spatially imaged iris plane screen.

gradually increases. In this system, the person perceives that the object is at the position of depth S due to the motion parallax.

This principle enables the intermediate viewpoint between the two viewpoints to be continuously interpolated and enables smooth viewpoint switching in the motion parallax even with a small number of projectors. We verified this principle earlier with a prototype based on conventional optical lenses and thus demonstrated the viability of a natural glassless 3D display harnessing both motion parallax and binocular parallax [3]. The only problem with the earlier prototype is that we were unable to easily increase the size of the display area or widen the viewing area due to physical constraints of the optical lens system.

2.2 Spatially imaged iris plane screen

We developed a spatially imaged iris plane screen that presents an image of the iris plane of the projector in the air. This screen consists of a reflective layer, a UV (ultraviolet) polymerized Fresnel lens layer, and an anisotropic diffusion layer, as shown in **Fig. 4**. This screen forms an iris plane at position d that

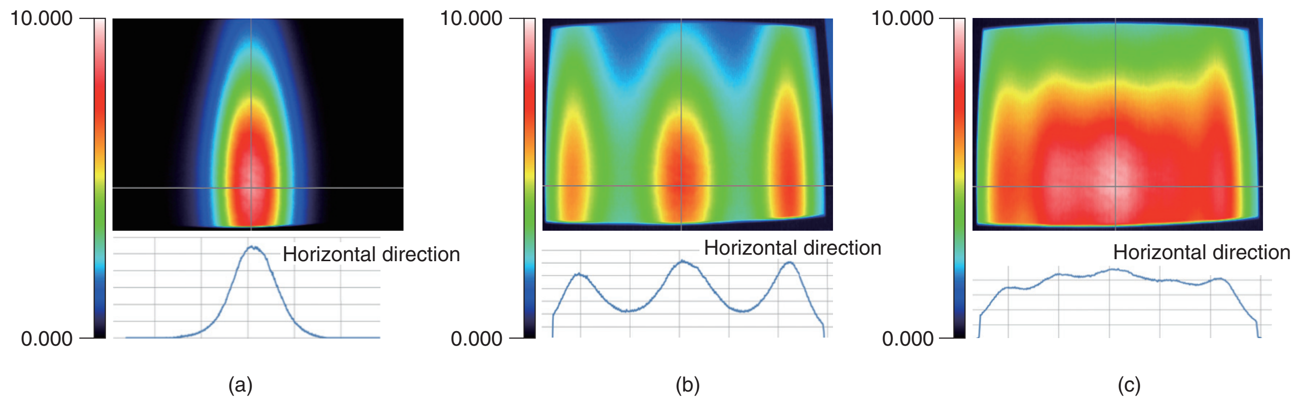


Fig. 5.   Luminance distribution of obtained spatially imaged iris plane.

satisfies 1 / *f* = 1 / *a* + 1 / *d*, assuming that the projection distance on the front surface of the screen is *a* and the focal length of the Fresnel lens is *f*.

When the iris plane of the projector is sufficiently small, the luminance distribution around the line of the projection plane can be controlled by the characteristics of the anisotropic diffusion layer. When the diffusion characteristics are narrowed, the light from the projector gathers in a narrow range. Therefore, it is considered that the smooth luminance distribution characteristics accompanying the movement of the viewpoint necessary for linear blending can be achieved. In addition, this screen is thin and flexible, and because it is a front projection type, it can easily be installed in a smaller space than that required for the conventional rear projection type.

We made a 50-inch prototype of the spatially imaged iris plane screen and conducted experiments to measure the luminance distribution and to project multi-viewpoint images.

3. Experimental assessment

We report here the results of the above experiments.

3.1 Luminance distribution on the new optical screen

We measured the luminance distribution of the iris plane imaged by the optical screen. The luminance distribution characteristics of the iris plane for a single projector are shown in **Fig. 5(a)**. One can see that as a result of the anisotropic diffusion layer with large diffusion only in the vertical direction, luminance is only uniformly spread in the vertical direction, while diffusion in the lateral direction is Gaussian. The slope reveals a linear luminance change as the luminance gradually decreases as it moves away from the center of the projection plane.

We measured how five iris planes overlapped in the region of half-value luminosity when five projectors were deployed. The luminance distribution of the spatially imaged iris plane for three projectors is shown in **Fig. 5(b)**, and the luminance distribution of the iris plane for five projectors is shown in **Fig. 5(c)**. One can see that when images are projected from multiple projectors, linear blending works properly in accordance with the luminance ratio, and iris plane luminance from multiple overlapping projectors becomes constant.

3.2 Prototype system and projection evaluation

We set up a multi-view image projection system with five projectors (**Fig. 6**) and then evaluated projected images using the prototype. Photographs taken of the center of the iris plane from five viewpoints and four mid-points between the viewpoints are shown in **Fig. 7**. It can be seen from these photographs that different images are observed depending on the position of the viewpoint, and that the intermediate viewpoint is also interpolated. Therefore, we confirmed that the spatially imaged iris plane screen provides natural stereoscopic viewing with motion parallax.

3.3 Discussion

We explain here how we evaluated the relationship between projector spacing and reproducible depth in the proposed system. Let *A* be the fusion limit angle, *d* the distance between the screen and the iris plane, *S* the perceived depth distance, and *t* the interval between the centers of the iris plane.

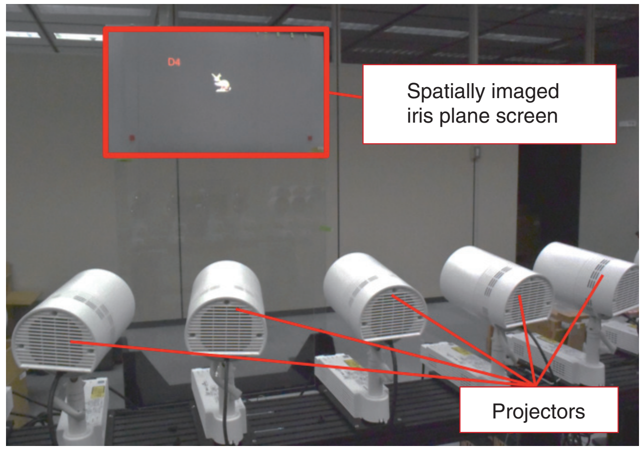


Fig. 6. Experimental system of multi-view projection using five projectors.

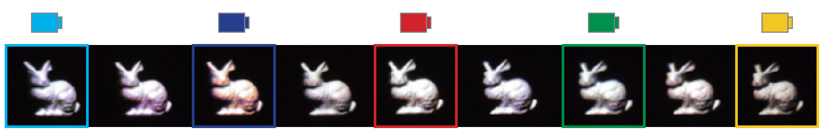


Fig. 7. Observation results from multiple viewpoints (five projectors).

Depth S is calculated by the interval t between the viewpoints as follows:

$$S = (d * v) / (t - v)$$
$$v = d * \tan (A) \quad (1)$$

The pop-out distance S' on the near side is calculated as follows:

$$S' = d * v / (t + v) \quad (2)$$

Therefore, if the fusion limit angle A is 8 minutes [6], and the distance between the screen and iris plane d is 2000 mm, the reproducible depth interval S is about 39 mm, and the pop-out distance S' is about 37 mm. Narrowing the interval between adjacent projectors makes it possible to increase the repeatable depth spacing.

4. Future work

We successfully developed a glassless 3D display screen adapted to viewpoint movement that dramatically reduces the number of required projectors. We demonstrated through assessment of projections that with our approach, we can effectively slash the number of projectors from one-third to one-tenth the number required in the conventional glassless 3D scheme accompanied by motion parallax using multi-view images.

The fact that our screen uses only a few video sources and does not require software processing since it is under optical luminance control makes it especially attractive for remote communications applications and for live broadcasts requiring real-time processing. Moreover, as a Fresnel lens-based light-concentrating device, the screen can also be implemented as a very thin ultrahigh luminance front projection screen. This makes it highly portable, easy to install and set up, and ideal for outdoor 3D signage and a wide range of other potential applications.

References

[1] Y. Takaki, "High-density Directional Display for Generating Natural Three-dimensional Images," Proceedings of the IEEE, Vol. 94, No. 3, pp. 654–663, 2006.

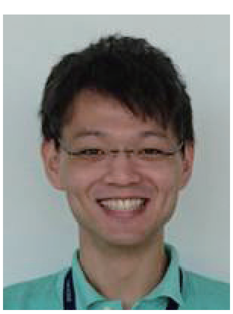
[2] A. Jones, J. Unger, K. Nagano, J. Busch, X. Yu, H-Y Peng, O. Alexander, M. Bolas, and P. Debevec, "An Automultiscopic Projector Array for Interactive Digital Humans," Proc. of ACM SIGGRAPH 2015 Emerging Technologies, Article No. 6, Los Angeles, CA, USA, Aug. 2015.

[3] T. Kawakami, M. Date, M. Sasai, and H. Takada, "Optical Linear Blending of Viewing Zones Using Convolution of Iris for Smooth Motion Parallax Autostereoscopic 3D Display," J. of Display Technology, Vol. 12, No. 2, pp. 143–152, 2016.

[4] M. Makiguchi, T. Kawakami, M. Sasai, and H. Takada, "Smooth Motion Parallax Glassless 3D Screen System Using Linear Blending of Viewing Zones and Spatially Imaged Iris Plane," SID 2017 Digest of Technical Papers, 61-2, Vol. 48, No. l, pp. 903–906, Los Angeles, CA, USA, May 2017.

[5] H. Takada, M. Makiguchi, T. Kawakami, and M. Sasai, "Smooth Motion Parallax Glassless 3D Screen System Using Linear Blending of Viewing Zone and Spatially Imaged Iris Plane," ITE Tech. Rep., IDY2017-34, Vol. 41, No. 24, pp. 13–16, 2017.

[6] H. Takada, S. Suyama, and M. Date, "Evaluation of the Fusional Limit between the Front and Rear Images in Depth-fused 3-D Visual Illusion," IEICE Trans. Electron., Vol. E89-C, No. 3, pp. 429–433, 2006.



Motohiro Makiguchi
Researcher, Natural Communication Project, NTT Service Evolution Laboratories.

He received a Master of Information Technology from the Graduate School of Information Science and Technology, Hokkaido University, in 2012. He joined NTT Service Evolution Laboratories in 2012. He has been researching 3D display technology and virtual reality technology.

Hideaki Takada
Senior Research Engineer, Supervisor, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E. in information and communication technology from Tokai University, Shizuoka, in 1995, an M.E. in information systems from the University of Electro-Communications, Tokyo, in 1997, and a D.S. in global information and telecommunication studies from Waseda University, Tokyo, in 2007. He joined NTT Integrated Information & Energy Systems Laboratories in 1997. He has been researching 3D visual perception and 3D display technology. He received Best Paper Awards from the 3D Image Conference in 2001 and 2005, the Young Researcher's Award from the Virtual Reality Society of Japan in 2002, the Achievement Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2003, a Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology of Japan in 2006, and Technical Committee Prize Paper Awards from the Institute of Electrical and Electronics Engineers (IEEE) Industry Applications Society in 2014 and 2016. He is a member of IEEE and the Institute of Image Information and Television Engineers.

Real-time Extraction of Objects with Arbitrary Backgrounds

Hidenobu Nagata, Hiromu Miyashita, Hirokazu Kakinuma, and Mariko Yamaguchi

Abstract

With the development of Kirari! immersive telepresence technology, we strive to achieve a high sense of realism, presenting objects that are at a remote location as though they were right before the viewer. One Kirari! function making it possible to achieve this type of expression is arbitrary background real-time object extraction. This involves extracting objects from backgrounds such as a playing field or theater stage without requiring studio equipment such as a green screen. This article gives an overview of this technology and introduces further efforts toward increasing its speed, robustness, and detail.

Keywords: object extraction, high realism, immersive telepresence

1. Introduction

The objective of Kirari! immersive telepresence technology is to transmit an event happening at a venue to remote locations in near real time, presenting athletes or performers to viewers with extremely high realism so that they appear to be performing right in front of viewers. *Objects* such as athletes or performers are extracted from the video and transmitted in real time to the remote location, together with other information surrounding the objects and comprising the venue space such as the audience and other background video, sounds, and light levels. Then the space at the remote location is reconstructed from the data.

The objects are represented in life size, as though the performers themselves were actually there. To achieve this, the video itself must be very clear in order to give the impression that the athlete or actor is actually standing in front of the viewer, and the region containing the objects must be extracted and displayed very accurately. At NTT, we believe we can provide experiences that are more exciting and have greater impact if the audience feels the performance at the remote location is happening *right now* in front of them, rather than being prerecorded and processed video. For this reason, processing needs to be done in real time.

Usually in situations requiring real-time object extraction, a special photographic technique called chroma-keying is used. For chroma-keying, a screen in a single color such as blue or green is placed behind the object so that the difference in color between the object and the background screen can be used to extract the object. Our goal with the Kirari! immersive telepresence technology was to establish technology able to extract the object accurately and in real time from video taken of the space surrounding the object, whether it be a sports field or stage, rather than using a technique such as chroma-keying. This is the arbitrary background real-time object extraction that we are working on.

2. Overview of object extraction system

Work on the arbitrary background real-time object extraction technology is advancing in three directions: to increase speed, robustness, and detail. We discuss these directions below, but first we give an overview of the overall process used by the object extraction system and explain the context for the work in each of these directions.

An overview of the object extraction process is shown in **Fig. 1** [1]. The current object extraction

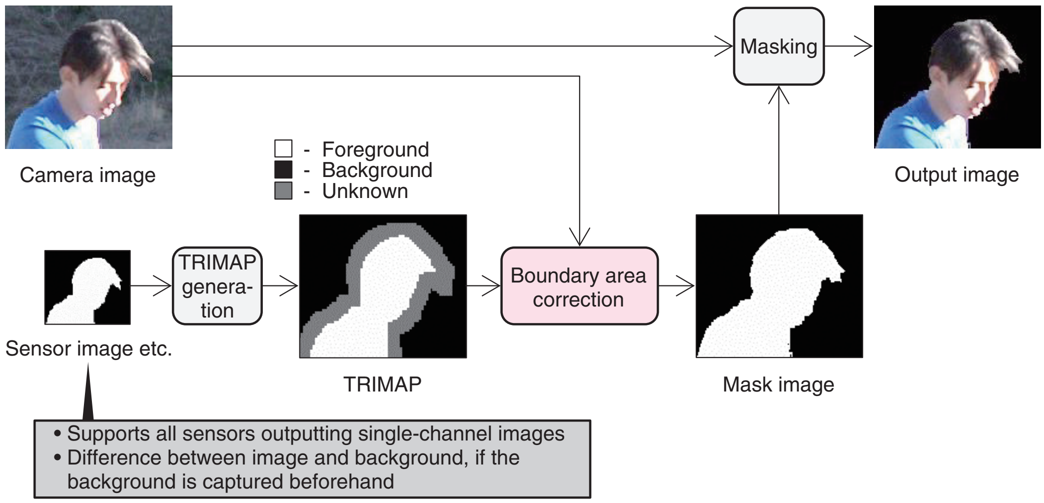


Fig. 1. Object extraction process.

system is based on the assumed use of sensors in addition to the imaging camera to capture the object. This is done to eliminate constraints on object extraction so that the range of applications can be expanded as much as possible in the future. As research on object extraction technology continues, further increases in the robustness of this process are being achieved.

Processing can be divided broadly into two stages. In the first stage, a sensing device and the background image (only if it can be captured ahead of time) are used to roughly identify the object area. In this stage, data (a trimap) are generated, and pixels are labeled in one of three categories: foreground, background, or unknown. Then, pixels in the trimap in the region labeled unknown (for which the foreground and background could not be distinguished during the rough identification stage and is assumed to contain the object boundary) are rapidly further classified as foreground and background, and corrections are applied to the boundary. Research continues to be done on this process, which is leading to increases in speed.

Finally, some objects targeted for extraction have an extremely complex boundary, such as with some traditional theater costumes. For these objects, the boundaries must be represented in fine detail to avoid degrading the sense of realism. In such cases, pixels in the unknown region are classified as foreground or background, but they are also assigned a transparency value called an α (alpha) value, which provides a more natural-looking boundary. The research and development (R&D) in this area is contributing to improvements that make it possible to achieve increased detail.

3. Three initiatives for technical development

Efforts are underway to improve the extraction technology. We describe here three improvements we hope to achieve and the initiatives being carried out to achieve them.

3.1 Initiatives for increasing speed

The boundary correction process used in the arbitrary boundary real-time object extraction technology is based on the assumption that the object will be extracted in detail, even when using low-resolution sensor images, and that the boundaries of the object area will be corrected using color data. The boundary correction process is based on a nearest-neighbor method, in which for each pixel labeled unknown in the trimap, the nearest (most similar) neighbor is found based on color data and coordinates, and the label of that pixel is copied to the pixel in question.

However, correcting this boundary is computationally intensive and becomes a bottleneck in producing the output video. Meanwhile, devices for capturing and displaying 4K/60 fps video are becoming more common, and there is increasing demand for Kirari! to support 4K/60 fps to provide an even higher sense of realism. However, processing high-resolution video captured using such equipment requires even more computation to perform boundary correction.

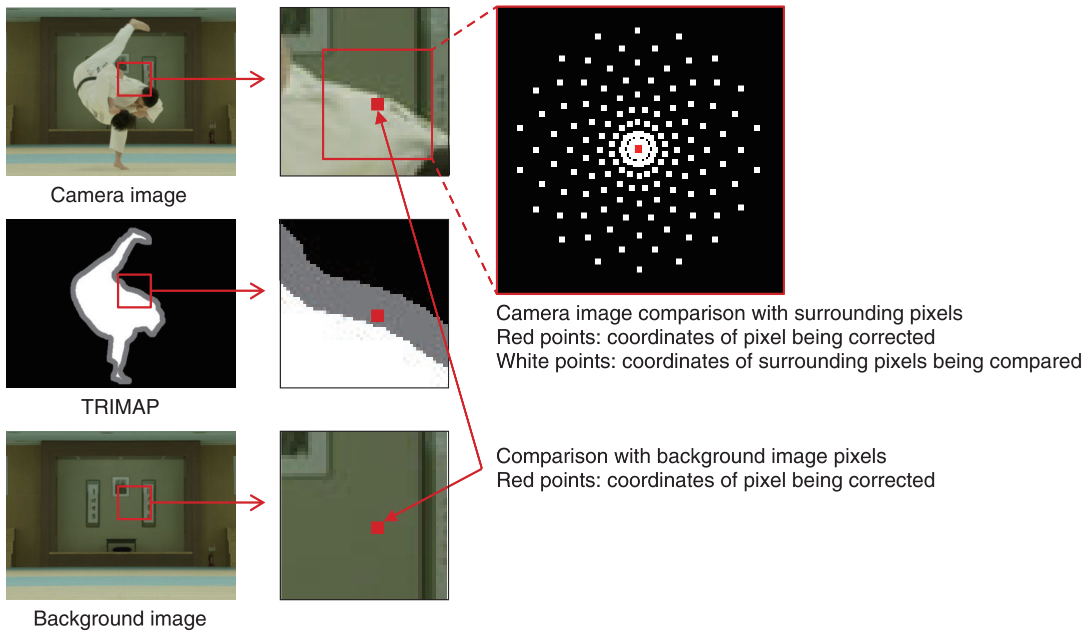


Fig. 2. Boundary correction.

As such, we set an objective to achieve real-time object extraction from 4K/60 fps video and are working to increase the speed of the boundary correction process, focusing in particular on reducing the amount of computation while also increasing the accuracy of extraction.

The boundary correction process has already been parallelized, but more technically, we can reduce the amount of computation by eliminating as many unnecessary comparisons as possible. To do so, we considered ways of searching for pixels that are *close enough* in terms of color data and distance between coordinates, rather than *the closest* in order to reduce computation while still obtaining a correct boundary.

When corrections are applied to the boundary, the pixel being corrected is compared with neighboring pixels. We adjusted this process so that comparisons with neighboring pixels are done more densely near the pixel and more sparsely further away. Comparisons are also done in order of increasing distance, meaning that the pixels that are nearest are compared first (**Fig. 2**). We also established thresholds and added a mechanism that stops doing comparisons and applies the label immediately if the distance derived from the color and coordinate data is below this threshold. In this way, *close enough* pixels are found quickly, and any further comparison operations are skipped.

However, the high resolution of 4K/60 fps video makes any error in extraction of the object noticeable, so higher accuracy than before is required. A particular weakness when generating the trimap with earlier techniques was that any error in labeling pixels as foreground or background was propagated to neighboring unknown pixels.

Therefore, we proposed the following method for obtaining the background image. For each unknown pixel, before performing any comparisons, we take the pixels at the correction coordinates in the input and background images. If the color information is close enough, we set the label of the pixel being corrected to background and avoid any comparison operations. This avoids propagation of any errors in background labels and makes it possible to obtain the correct object region [2].

3.2 Initiatives for increasing robustness

Objects to be extracted are not always in static environments or situations such as a studio that can be easily controlled. For example, the lighting and shadows behind the object can change during a sporting competition, and naturally, there are scenarios with spectators and other objects moving in the background. We are increasing the robustness of our technology to realize real-time object extraction that can also handle these sorts of situations.

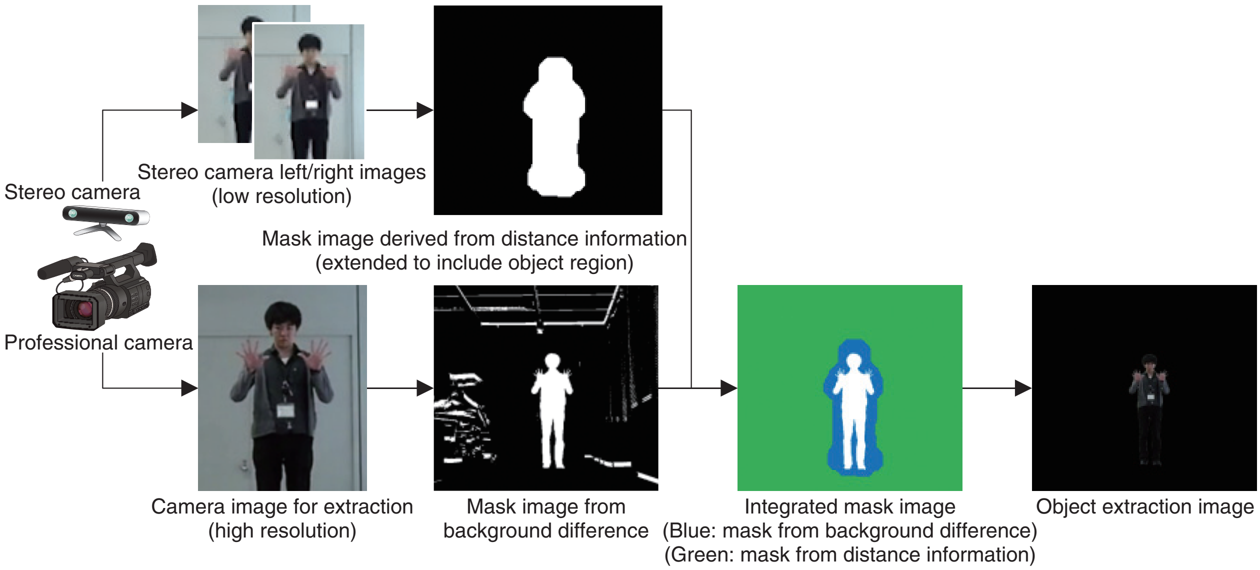


Fig. 3. Object extraction system incorporating a stereo camera.

At the NTT R&D Forum 2017 held in February, we built a system able to extract a specified object from full high-definition video in real time (30 fps or better) by combining video from a professional camera with partial depth information obtained using an off-the-shelf stereo camera (**Fig. 3**).

With this system, the system operator first selects an object to be extracted and specifies a rough range of distances for the object based on object distance information obtained using the stereo camera. Then, regions outside the specified depth range and regions with no change relative to the background are masked. This improves robustness for cases where the background changes, over just comparison with the background. This enabled the Kirari! system to extract one researcher from among several people in a meeting room, transmit the image, and project it onto the stage at the NTT R&D Forum 2017 in a remote location, where the researcher was able to interact with the moderator there.

However, although this system was able to roughly mask out undesired objects based on distance, it did not have sufficient accuracy to accurately separate the object border from the background (i.e., accuracy to correctly separate the object boundary when there are moving objects in the background from the camera's perspective). The system used an off-the-shelf stereo camera, so the mask image generated using distance data from the stereo camera was of a lower frame rate and had a different angle of view than the main professional camera image, from which the object was extracted. This resulted in frame mismatch between the stereo and main cameras when the object being extracted moved quickly or over a wide range, which decreased accuracy in extracting the object.

In the past, we have studied combinations of a camera used with various types of sensors such as time-of-flight and thermal sensors, but we found that images taken by all of these off-the-shelf sensors—just as with the stereo camera—differed from those taken by the main camera in resolution, field of view, or frame timing, and it was extremely difficult to calibrate and synchronize them perfectly, either spatially or temporally. The accuracy of measurements taken by sensor devices was also greatly affected by factors such as type of clothing, lighting, and measurable range. Making the object extraction system more versatile is therefore an issue to be addressed, and we intend to further improve the system to make it more robust in the future.

3.3 Initiatives for increasing detail

Kirari! is also desired for use in entertainment, for events such as concerts and *kabuki* theater. As such, we can envision scenarios when we would like to extract performers wearing complex and finely detailed costumes in traditional theater or a live concert. In such situations, the boundaries of extremely detailed hair styles or the textures of translucent costumes must be reproduced faithfully. This requires a

Fig. 4. Application of minimal matting process.

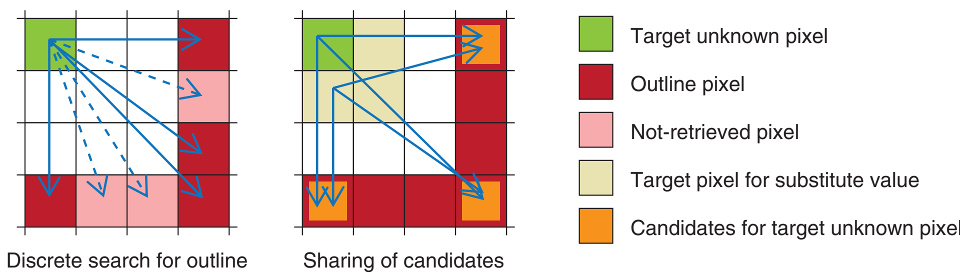


Fig. 5. Approximation in the matting process.

method for reproducing these complex and detailed boundaries that is more effective than the method of classifying pixels and applying corrections to the boundary.

We have studied a method that assigns a transparency value to pixels marked as unknown in the trimap in order to present extracted results naturally and in detail, even for objects with complex boundaries such as these. The method is called matting, and it is well known for naturally blending an extracted object into another image (background). Various algorithms have been proposed for matting, but algorithms to extract objects without the use of a green screen as in chroma-keying require large amounts of computing time and are difficult to use for real-time object extraction.

Because not all objects have complex boundaries, we only apply matting when necessary by classifying images according to edge intensity when the contrast with the background is low or the boundary region is very complex, as with some hairstyles [3]. For boundary regions where the object and boundary can be clearly distinguished, we apply boundary correction as in our earlier method (**Fig. 4**). This enables us to minimize the area that requires heavy computation, reducing the overall computation time.

We also introduced some approximation into the matting process. For matting, the transparency value applied to the pixel is derived from the surrounding pixels (candidate points), but candidate points are computed for each pixel handled, which increases the computation time. We therefore divided the matting region into small subregions and computed candidate points for each one, reducing the number of computations that need to be done. Note that within the small subregions, neighboring pixels have similar brightness and color, so computing transparency with this type of approximation enables extraction to be done without greatly degrading the quality of the boundary region (**Fig. 5**). We also introduced parallel processing using a graphics processing unit (GPU), which increased the processing speed by a factor of about 30 compared to our earlier implementation. The resulting process was used in the highlights section of the Kabuki Virtual Theatre demonstration held in Kumamoto, Japan, in March 2017. We will continue to study techniques for increasing speed in the future.

4. Future prospects

We have introduced efforts to improve our arbitrary background real-time object extraction technology, a core technology of Kirari!, in the three areas of speed, robustness, and detail. Regarding speed, we have achieved a 4K/60 fps processing speed while increasing accuracy. We have improved robustness with respect to background content and capture environment using an off-the-shelf stereo camera to measure depth information. To increase detail, we produced an algorithm for processing complex boundaries that also minimizes any loss of speed. To this point, object extraction R&D has been advancing separately in the three approaches we have introduced in this article, but going forward, we plan to promote integration of this work, expansion of its applications through practical demonstration opportunities, and efforts toward practical application.

References

[1] H. Miyashita, K. Takeuchi, M. Yamaguchi, H. Nagata, and A. Ono, "Fast and Accurate Image Segmentation Technique Using Multiple Sensors," IEICE Tech. Rep., Vol. 116, No. 73, pp. 17–22, 2016.

[2] H. Miyashita, K. Takeuchi, H. Nagata, and A. Ono, "Fast Image Segmentation for 4K Real-time Video Streaming," IEICE Tech. Rep., Vol. 117, No. 73, pp. 189–190, 2017.

[3] M. Yamaguchi, H. Nagata, and A. Ono, "Adaptive Matting for Fast, Accurate Object Extraction," Proc. of 2017 IEICE General Conference, D-11-29, Nagoya, Aichi, Japan, Mar. 2017.

Hidenobu Nagata

Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received an M.E. in systems and information engineering from Hokkaido University in 2001. He joined NTT in 2001 and studied video technologies including video indexing and automatic summarization. From 2008 to 2014, he worked at NTT Electronics and developed transcoders and embedded audio Internet protocol (IP). He is currently researching ultra-realistic communication technology including the immersive telepresence technology called "Kirari!".

Hirokazu Kakinuma

Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received an M.E. in information science from the Graduate School of Advanced Integration Science, Chiba University, in 2010. He joined NTT in 2010 and studied database management systems for IP television (TV). From 2013 to 2016, he worked at NTT Plala, where he developed set-top box web applications and released the Hikari TV 4K service for 4K digital TV. He moved to NTT in 2016 and is currently researching real-time image segmentation technology based on machine learning.

Hiromu Miyashita

Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E. and M.E. from Keio University, Kanagawa, in 2008 and 2010. He joined NTT Service Evolution Laboratories in 2010, where he worked on human computer interaction, image and video processing, and ultrahigh-presence telecommunication services.

Mariko Yamaguchi

Natural Communication Project, NTT Service Evolution Laboratories.

She received an M.E. from the Department of Computational Intelligence and System Science, Tokyo Institute of Technology Interdisciplinary Graduate School of Science and Engineering in 2015. She joined NTT in 2015 and is currently working on image processing technology for a virtual reality system.

Real-time Moving Object Detection Technology and Trial of Stone Location Delivery at a Curling Venue

Tetsuro Tokunaga, Yoshihide Tonomura, Jun Shimamura, Fumihisa Hamamura, and Hiroaki Kakimoto

Abstract

NTT Service Evolution Laboratories is researching and developing tracking technology combined with video understanding as part of efforts to develop advanced video services. This technology features both high recognition performance and a short processing time and shows promise for application to public event viewings and stadium solutions. This article introduces a trial conducted at the 2017 Sapporo Asian Winter Games in which this technology was applied to deliver stone location information to spectators at a curling venue.

Keywords: moving object detection, image processing, curling

1. Introduction

NTT Service Evolution Laboratories is researching and developing tracking technology for real-time recognition and tracking of objects in video with the objective of developing more advanced stadium solutions and public viewing of events. The aim here is to achieve tracking technology that has both a short processing time and high-accuracy object recognition to provide in real time meaningful information to spectators of competitive matches, audiences of theatrical performances, and other event spectators. We can envision various types of applications using tracking technology combined with a video understanding system, for example, recognizing and tracking an athlete or the ball during a sports match and overlaying information as in augmented reality, and recognizing and tracking the motion of an actor's hands and adding stage effects. We can expect new video services to be offered in this way.

In this article, we outline the real-time moving object detection technology that we developed targeting curling stones and describe the trial we conducted at the curling competition held at the 2017 Sapporo Asian Winter Games using a system incorporating this technology.

2. Demonstration experiments on providing new spectator experiences at Asian Winter Games

The City of Sapporo and NTT held demonstration experiments on providing new spectator experiences using advanced information and communication technology at the 2017 Sapporo Asian Winter Games held in Sapporo, Hokkaido, from February 19 to 26. The targeted sports were figure skating, ice hockey, and curling, and these demonstration experiments used an official smartphone application (app) for the 2017 Sapporo Asian Winter Games to showcase new spectator services that could not be experienced by the traditional means of watching television or observing the sports at the venue.

When viewers watch a curling match, their ability to accurately understand the stone's current location is tied to having an accurate understanding of the state of the match. However, since it is impossible to obtain an overhead view of the ice-based *curling sheet* (playing surface) and *house* (the circular target

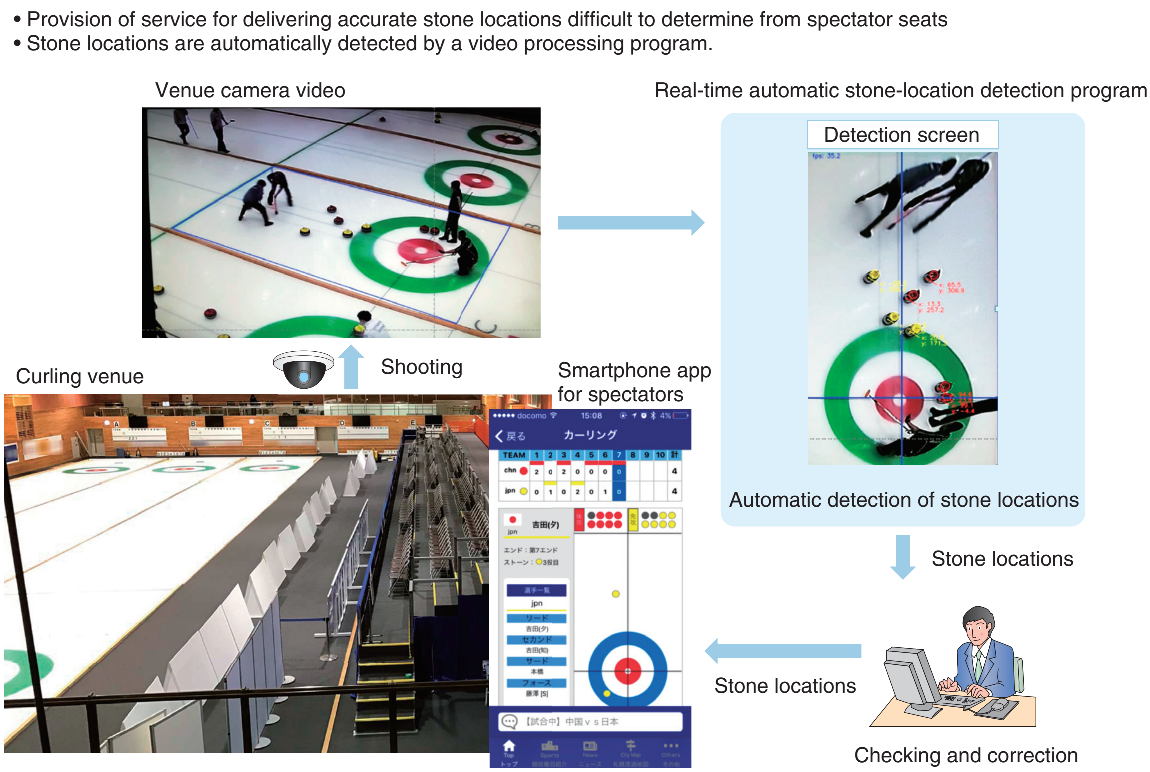


Fig. 1. Demonstration experiment (curling) on providing a new spectator experience at 2017 Sapporo Asian Winter Games.

on the sheet) from spectator seats inside the venue, it has been difficult to accurately grasp the locations of the stones. To address this problem, we aimed to provide a new spectator experience service enabling a spectator inside the curling venue to grasp the state of the current match by delivering stone location information to the smartphones of spectators.

To deliver stone location information that is useful in understanding the state of the match, the stone location must be obtained during the match as the match progresses. We therefore developed technology for detecting in real time the location of a stone as a moving object and conducted a trial on the use of moving object detection technology to deliver stone location information to spectators at a curling venue (**Fig. 1**).

3. Real-time moving object detection technology targeting curling stones

To achieve low-cost detection of the locations of all stones present within the effective area between the hog line and back line without making modifications to players or their equipment, we decided to make use of existing venue stones and cameras. The camera above the sheet of a match in progress has a limited shooting range and cannot be controlled freely, so we used a camera above a sheet not currently in use and oriented it to overlook and shoot the effective area (**Fig. 2**).

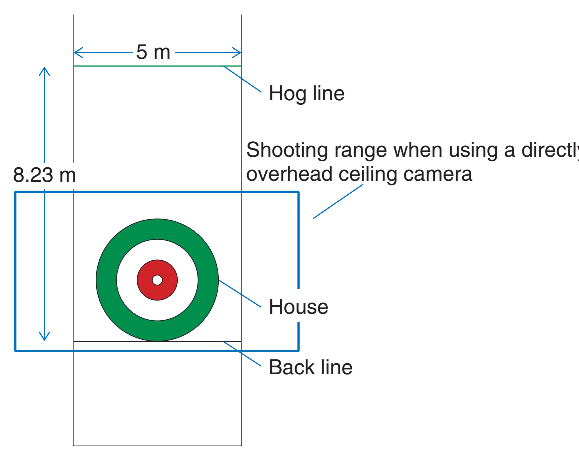


Fig. 2. Curling house area.

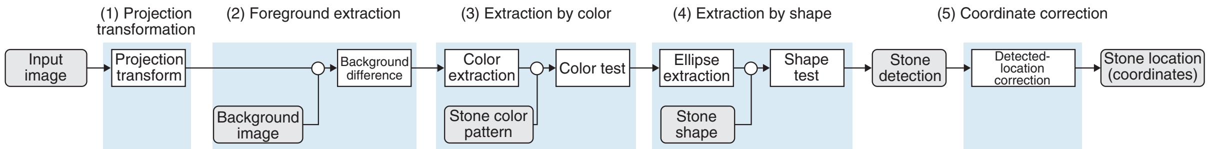


Fig. 3. Process flow per frame of real-time automatic stone-location detection program.

We developed technology for inputting video taken in this way and for detecting in real time the team color and location coordinates of each stone. The process flow for each video frame is shown in **Fig. 3** and described below:

(1) Projection transformation: A projection transformation is applied to the oblique view of the input image shown in Fig. 1 to achieve an overhead image. Subsequent processing is limited to the effective area after this projection transformation to reduce computational complexity.

(2) Foreground extraction: A background image less any stones or athletes is shot, the foreground is extracted by background difference processing, and the result is passed to the next process as stone candidate areas.

(3) Extraction by color: Only two colors—red and yellow—were used for the colors of stone surfaces in this trial, so two types of color patterns were recorded from stone images prepared beforehand. This process compares these red and yellow patterns with the stone candidate areas, extracts those areas with similar colors, and passes the result to the next process.

(4) Extraction by shape: The stone candidate areas input to this process would include athlete uniforms if their colors were reddish or yellowish. To therefore exclude such candidate areas and detect only stones, this process detects the outline of each input stone candidate area and uses its shape and size to judge whether that candidate area corresponds to a stone.

(5) Coordinate correction: The above projection transformation assumes a flat target, so any object having height as in the case of a stone surface will be transformed to a location offset from the actual location. This process corrects each detected location to estimate a more accurate location and finally outputs sets of stone colors and stone coordinates as program results for this frame.

4. Stone location delivery system for spectators at curling venue and trial

We constructed a stone location delivery system based on real-time moving object detection technology that we developed. This system delivers stone location information to a smartphone app every time a throw is completed, enabling spectators at a curling venue to understand where the stones in a match are currently located. In this trial, a function for recognizing the time that a throw is completed (the time at which all stones are stationary) was not implemented as part of the program but simply achieved by having the operator determine that time visually.

The basic configuration of this system for one curling sheet is shown in **Fig. 4** and summarized below.

As previously described, video signals from existing camera facilities in the curling venue are distributed and used as input video to the proposed system. In curling, however, the stone-throwing direction changes every *end**, so the effective area enclosed by the input image must be switched for each end. The system therefore installs a switcher so that the video input to the video analysis personal computer (PC) can be selected from the video output from two cameras.

Now, once the operator visually determines that a throw has been completed and operates the system accordingly, the video analysis PC detects the location of that stone from the input video using real-time moving object detection technology and passes that result to the delivery control PC.

Next, the operator at the delivery control PC visually checks the received stone location information for errors, corrects errors if any exist, and uploads the result to the delivery server. The smartphone app

* End: In curling, one round of play by each team in a game. One end is completed when each team throws 8 stones, for a total of 16 stones.

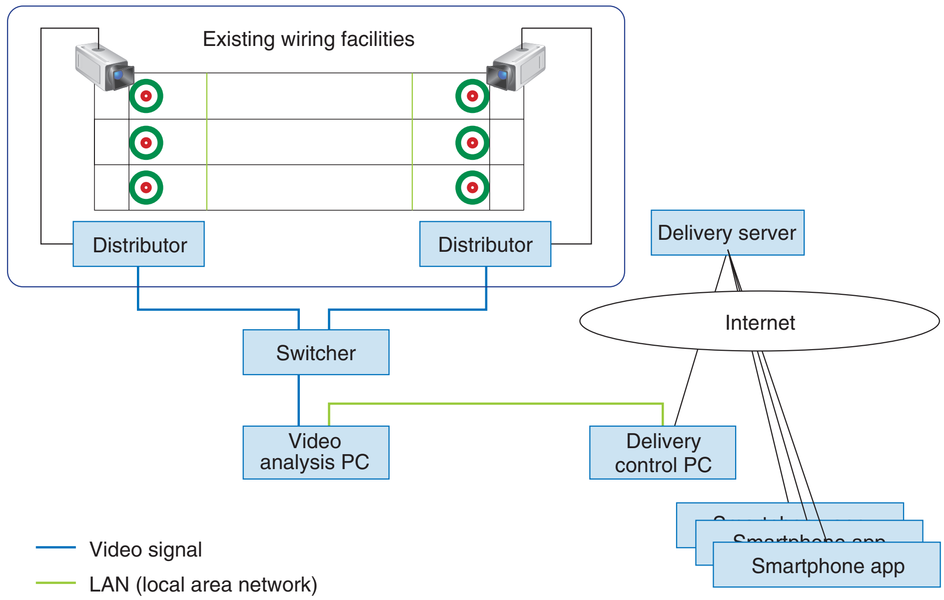


Fig. 4. Configuration of curling-stone location delivery system.

accesses the delivery server every few seconds. If the app detects an update on the stone location information at that time, it obtains that information and updates its display content.

This system trial was conducted during the curling competition held February 18–24. The system operated without problems during this period and received a favorable response from users and curling associations.

5. Accuracy of stone location detection results in trial

We evaluated the accuracy of the stone location detection results using video of 21 matches of the men's and women's curling competitions at the 2017 Sapporo Asian Winter Games. The stone location after being visually checked and corrected if necessary by the operator of the delivery control PC was taken to be correct location information.

For each match, the number of stones corrected by the operator of the delivery control PC, the total number of stones placed in the effective area and targeted for location detection, and the stone correction rate are listed in **Table 1**. The stone correction rate per match fell in the range from 1.12% to 25.98%, resulting in a value of 11.12% for all matches. This means that the probability of having no correction needed by the operator of the delivery control PC was approximately 90%.

Plots of corrected stone locations on the curling sheet are shown in **Fig. 5** by stone color. A comparison of the two sets of plots indicates that the number of corrections is higher for red stones and that many of these red-stone corrections appear on the red circle of the house. We surmise that the close resemblance of the house color and the red stone color led to errors in detection.

6. Future development

In this article, we described real-time moving object detection technology targeting curling stones, a stone location delivery system incorporating this technology, and a system trial conducted at the 2017 Sapporo Asian Winter Games. In the trial, the stone locations were detected with an accuracy rate of about 90%. We were able to provide a new spectator experience service that delivered stone location information to spectators at the curling venue without any problems during the period of these games.

Going forward, we plan to continue our research and development efforts to enhance tracking accuracy using moving-object locus information and to expand the number of recognizable moving objects. In this way, we seek to provide new video services featuring, for example, new ways of viewing sports and new types of staging in entertainment.

Table 1. Stone correction rate per match.

Match ID	Number of stones corrected per match	Number of stones placed in effective area per match	Correction rate (%)
1	43	416	10.34
2	43	606	7.10
3	69	582	11.86
4	18	669	2.69
5	95	552	17.21
6	63	570	11.05
7	23	471	4.88
8	38	492	7.72
9	56	334	16.77
10	20	560	3.57
11	90	568	15.85
12	35	551	6.35
13	65	583	11.15
14	88	594	14.81
15	28	383	7.31
16	123	547	22.49
17	157	701	22.40
18	5	447	1.12
19	27	546	4.95
20	153	589	25.98
21	31	657	4.72
Total	1270	11,418	11.12

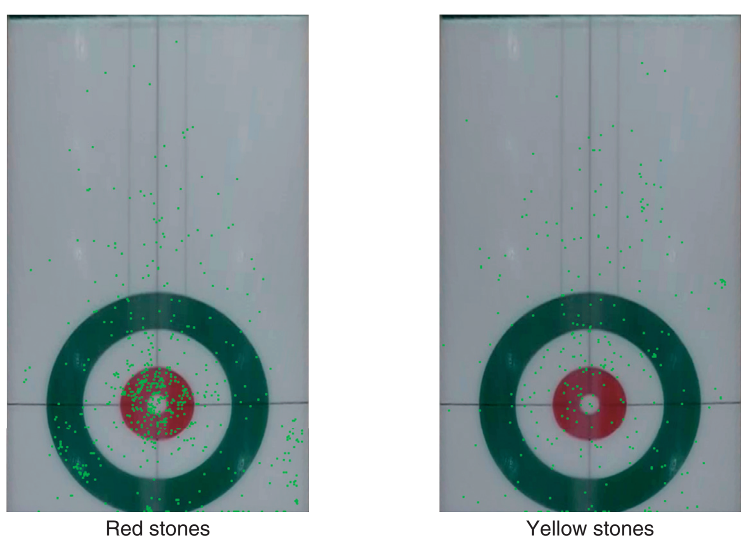


Fig. 5. Corrected stone locations (by stone color).

Tetsuro Tokunaga

Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received a Bachelor of Liberal Arts and Master of Informatics from Yokohama National University, Kanagawa, in 2002 and 2004. He joined NTT Network Service Systems Laboratories in 2004. He was with NTT Resonant from 2008 to 2012. His current research interests include web technology, computer vision, and design thinking. He was a co-developer on a team that participated in 2003 in the youth division in the Exploratory IT Human Resources Project (The MITOH Program) initiated by the Information-technology Promotion Agency of Japan. He is a member of the Information Processing Society of Japan and the Database Society of Japan.

Yoshihide Tonomura

Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received an M.S. in electronics engineering from Nagaoka University of Technology, Niigata, in 2004 and a Ph.D. from Tokyo Metropolitan University in 2010. Since joining NTT, he has been researching and developing multimedia data transmission technologies. From 2011 to 2012, he was a visiting scientist at MIT Media Lab, Massachusetts, USA. He received the SUEMATSU-Yasuharu Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2016 and the 26th TAF TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2011.

Jun Shimamura

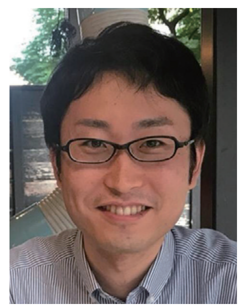
Senior Research Engineer, Planning Section, NTT Service Evolution Laboratories.

He received a B.E. in engineering science from Osaka University in 1998 and an M.E. and Ph.D. from Nara Institute of Science and Technology in 2000 and 2006. He joined NTT Cyber Space Laboratories in 2000. He was with NTT Communications from 2006 to 2009 and NTT Media Intelligence Laboratories from 2010 to 2015. His research interests include computer vision and mixed reality. He received the 2005 TAF TELECOM System Technology Award from the Telecommunications Advancement Foundation. He is a member of IEICE and the Institute of Image Information and Television Engineers.

Fumihisa Hamamura

Associate Manager, Strategic Business Development Division, NTT.

He received a B.A. in environment and information studies from Keio University, Kanagawa, in 2005 and joined the NTT Group the same year. Since 2012, he has been engaged in new business and service development with ICT in the healthcare and sports market.



Hiroaki Kakimoto

Strategic Business Development Division, NTT.

He received a B.E. and M.E. in computer science from Nagoya Institute of Technology, Aichi, in 2008 and 2010. He joined NTT Communications in 2010 and was engaged in corporate sales and development of a live broadcasting service. He has been with NTT since 2015 and has been involved in business development and collaboration support for NTT Group companies in sportstech, healthtech, and other fields.

Initiatives Concerning Development of Applications Utilizing Blockchains

Atsushi Nakadaira, Shigenori Ohashi, Hiroki Watanabe, Shigeru Fujimura, Satoshi Sakuma, and Shingo Kinoshita

Abstract

The development of services utilizing blockchains is increasing. In this article, we explain issues related to the design of applications using blockchains that require further investigation and use cases that exploit the advantages of blockchains. The use cases focus on the distribution management of management targets, namely, data managed by blockchains. Development of a common-function module for supporting development of applications is also described.

Keywords: blockchain, bitcoin, distributed application

1. Introduction

Blockchain technology was announced by a person (or persons) going by the name of Satoshi Nakamoto, and it is known as a core technology of the cryptocurrency called bitcoins, which emerged in 2008 [1]. The advantage of blockchain technology is that it enables data to be managed in the manner of an autonomous distributed system that does away with a centralized organization; consequently, it is difficult to falsify data. Thanks to this feature, blockchain technology is drawing attention in regard to not only cryptocurrency but also applications such as management of banking systems, property, and rights. In addition, the application of this technology is expected to be extended to *sharing* services (i.e., services that do not depend on a specific management organization) and services utilizing the Internet of Things (IoT) [2].

With blockchain technology, there seems to be more focus on discussing the ideas and methods related to the mechanism of the technology itself than on discussing its necessity or how best to utilize it. As a result, there is a need to conduct trials and verifications concerning utilization of the technology for developing actual services and promoting business development. In this article, we first discuss the blockchain technology that is practically applicable to services. Then we describe elements that must be confirmed when developing applications utilizing blockchain technology. We also explain NTT's initiatives to efficiently develop applications. These initiatives essentially target the application layer.

First, services that would be applicable to blockchain technology were investigated. One highly anticipated application of blockchain technology that was identified was the concept of *smart property*. Smart property refers to assets that can be managed on a computer network in the manner of a cryptocurrency. These assets include both physical and digital assets. The former is managed in terms of usage rights of vehicles, ownership of land, stock certificates, and so on, and the latter is managed in terms of viewing rights of digital content, utilization rights of IoT data, and so forth. Registering information about assets (namely, an identifier (ID) that denotes a particular asset) in a blockchain and distributing that information in the manner of circulating currency is expected to enable the smart and flexible utilization of assets while also ensuring transparency. This kind of utilization is being investigated, but use cases other than cryptocurrency (e.g., services) that can inherently exploit the advantages of blockchains have not yet been clarified. Accordingly, we summarize in this article use cases that take advantage of the benefits of blockchains.

Additionally, we explain issues that should be considered when developing applications. Blockchain technology is being developed not only as a simple means of transferring value via a cryptocurrency but also as a system for executing digital contracts in an autonomous and decentralized manner. It is conceivable that such contracts—known as *smart contracts*—will be applied in a broader range of fields as blockchain technology becomes further developed.

In 2014, Ethereum, a new blockchain implementation that certifies program execution, was proposed [3], and it has been drawing attention ever since as a blockchain platform. In comparison with bitcoin blockchains, Ethereum makes it easier to manage extremely high-level data. However, it makes it more complicated to implement data structures of transactions, structures of blocks, and methods of managing a variable number of programs. To utilize the Ethereum platform, it is necessary to develop applications specialized for particular program-execution environments. At the same time, it is necessary to develop a framework for developing applications. Accordingly, we propose implementing common basic functions on the application layer and modularizing them in order to improve the efficiency of application development.

2. Utilization of blockchain technology

Blockchain technology has been proposed for use as the platform of various services. Such proposals are not limited to the transfer of value by cryptocurrency but also include management of assets and rights. Efforts concerning PoC (proof of concept) are also advancing. Various arguments have been made for use cases and applications exploiting the features of blockchains, but these arguments have not been put forth in a clear and orderly way. When data management in a distributed environment is considered in vague terms, the result of that consideration is likely to be that it is not only impossible to exploit the features of blockchain technology but also that it is better to use superior conventional methods. Accordingly, we express our ideas concerning key points of use cases utilizing the features of blockchains. The four conditions listed below must be considered when utilizing blockchains for managing data such as assets and digital content:

(1) Management is not performed by any single particular authority.
(2) Management targets can be reconfigured during the distribution process.
(3) The utilization form and distribution process after distribution starts are not determined (i.e., fluidity exits).
(4) Objective traceability of the distribution process is paramount.

First, it is presupposed that management is not performed only by any single particular authority. Alternately, it is presupposed that although management might be performed by a certain body taking a centralized role, it is necessary to ensure the *inspectability* of management status and information registration in regard to bodies and organizations that are independent of that central body.

One feature of blockchains is exploited that is relevant to the first condition, namely that authority is distributed among multiple bodies and managed accordingly. Under the first presupposition, it is conceivable that use cases will involve the repeated circulation* of information and digital content while the configuration of management targets is changed.

As for the second condition, namely, reconfiguring management targets during the distribution process, we use as an example the management target of video content. In this case, other images and sound are combined, and the combined content is distributed. In another example, if the management target is IoT-data groups, it is supposed that the data are smoothed, combined with other data, and further distributed alongside the results of analysis.

As for the third condition (concerning the distribution process), it is conceivable that blockchains can provide hitherto unavailable value not under the condition that the distribution process is set from the outset and managed under that status, but under the condition in which the distribution process cannot be specified. If the distribution process cannot be predetermined, two details are registered and managed: first, what kind of information is taken, how someone processes that information, and what kind of information is generated as a result; second, what rights and authority are granted to the people or organizations that took, processed, and generated the information.

Under the conditions described above, the place of origin and transitions of information being utilized can be confirmed by referring to the information registered in a blockchain. As a result, it is possible to

* Repeated circulation: The wide circulation of goods and rights by the ability to transfer them between users and changing owners in succession.

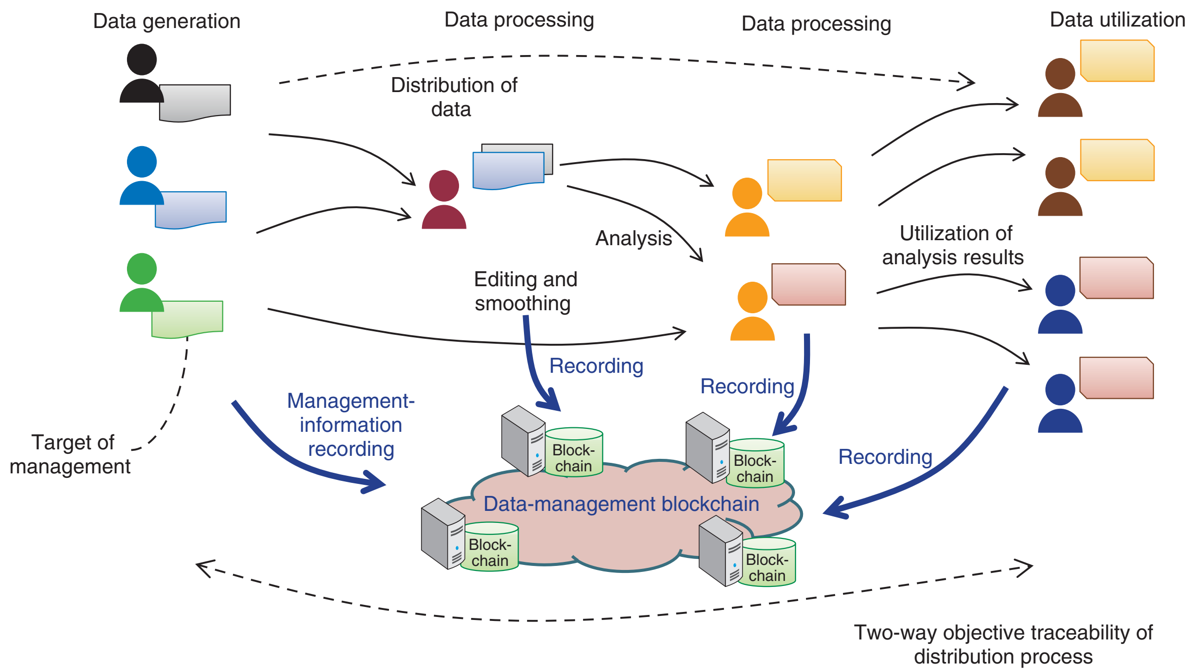


Fig. 1. Model of blockchain utilization service.

certify the authenticity and credibility of data or the existence of rights to use data as objective data. Since the data sources and the history of data changes are traceable, it is possible to evaluate the contributions to the value of data by dating back to the source of the data. Under those circumstances, to successfully utilize that traceability of data (and thereby assure objectivity), it is necessary to fully exploit the benefits of blockchains (**Fig. 1**).

As a result of utilizing blockchains in the manner of the use cases described above, information management (which has been conventionally governed by specific bodies) can be converted to decentralized co-management. Consequently, it is conceivable that on the business side, it is possible to eliminate intermediaries and thereby cut costs. However, it is considered unlikely that the important roles played in business by such intermediaries and wholesale firms will diminish, as a result of information matching, information arrangement, and other practices, and that the role itself will become unnecessary. In other words, it is conceivable that instead of such intermediaries simply disappearing, it is possible that the roles of intermediaries and wholesale businesses will become clarified through data traceability, and their value can be optimized accordingly.

3. Technical aspects required in designing applications

By applying blockchain technology in the manner described in the preceding section, it is possible to provide value that has hitherto been unavailable. However, in designing applications, the three technical aspects listed below must be considered:

(1) Assuring authenticity of registered data
(2) Devising methods for storing and managing appropriate data
(3) Assuring traceability of data

First, it is essential to assure the authenticity of registered data. In contrast to assuring authenticity of *coins* generated by a system in the manner of bitcoin, it is considered that when information is being managed, it is essential to affix extra information that confirms the authenticity of the registered information. A scheme for registering information while certifying the person who registered that information (by means of a digital signature) is thus required.

In accumulating data, it is necessary to manage methods of storing data as well as methods of accessing the stored data. Various data storage methods are conceivable, including a method in which both data and the data's management information are recorded in the blockchain, and a method that utilizes external storage for data with a huge volume such as video content, while the management information of that video content is managed via a blockchain. As mentioned previously, the latter scheme is referred to as *smart property*.

With smart property, it is essential to firmly link the main body of content stored in external blockchains and the management information within those blockchains. In managing access to stored data, it is necessary to manage not only access to the content itself but also access to the management information managed by blockchains. In a blockchain, each node can synchronize and hold the same ledger data as other nodes as well as browse the ledger content. Therefore, it is necessary to conceal the management information managed by blockchains when managing access to that information. We have been investigating information-access management by processing data via distributed applications as well as methods for encrypting and managing data for management information managed via blockchains.

Regarding the traceability of data, it is necessary to ensure that traceability of noteworthy data is not lost while enabling data accumulated in blockchains to fluctuate fluidly.

To meet these requirements, it is necessary to develop technologies for (i) registering data, (ii) storing and managing data, and (iii) tracing data.

In our focus on use cases so far, we have investigated content management by applying blockchains [4]. In particular, we have been developing elemental technologies by taking an application-layer approach for practically applying blockchains [5, 6].

We use the case of video images taken with a camera as an example to describe our data-registration technology for managing content. In this case, digital-content information listed by the content creator is recorded at the moment the content is created. The reliability of recorded information is validated using the digital-content information and information concerning applications (such as cameras used for creating content) for the verification. The application information is assumed to be recorded by the application developer in blockchains before it is distributed. Separating digital-content information and application information in this way makes it possible to integrate the reliability of individual digital content and attain overall reliability of shared applications.

- Digital-content information

1) Digital-content hash value
2) Information concerning creator and creation conditions
3) Digital signature corresponding to recorded information (i.e., signature for application used for creating digital content)

- Application information

1) Identification information (e.g., package name and application name, version, and developer's name)
2) Public key of application
3) Information about creation metadata (i.e., item information including user-setting availability)

The digital-content hash value and information concerning the creator and creation conditions of digital-content information are automatically configured by an application compatible with the proposed method as metadata on the terminal of the creator. For example, the creator registers the address of a wallet account used for issuing transactions. The digital-content hash value is used as an ID for uniquely indicating content. However, if the encoding method is changed, for example, to ensure that content can be identified, it is conceivable that video fingerprints can be registered alongside the hash values of content and that those fingerprints can be used to create IDs to replace the hash values.

Additionally, the date, time, and location are registered during the content creation. However, it is conceivable that multiple acquisition sources corresponding to that information exist. For example, if a time and date are registered, clock times from external NTP (network time protocol) servers as well as times on other devices are also registered. Accordingly, a means of registering those acquisition sources is implemented. A digital signature is assigned multiple purposes: specifying the application used for creating content; discriminating items that can be handled by people from those that cannot, in addition to discriminating application information; and confirming that information has not been rewritten.

Although the storage and management of data differ according to the target to be managed, the volume of video content itself is very large. As a result, a method for managing the content itself outside blockchains but managing content-management information and rights information via blockchains has been proposed [4].

A method exists for managing licenses to use content via blockchains; in concrete terms, access to externally managed content encrypted by a common key is managed by exchanging that key. There is also a method for exchanging an encrypted common key via blockchains. At the source (i.e., sender), the sender can encrypt the common key as the public key

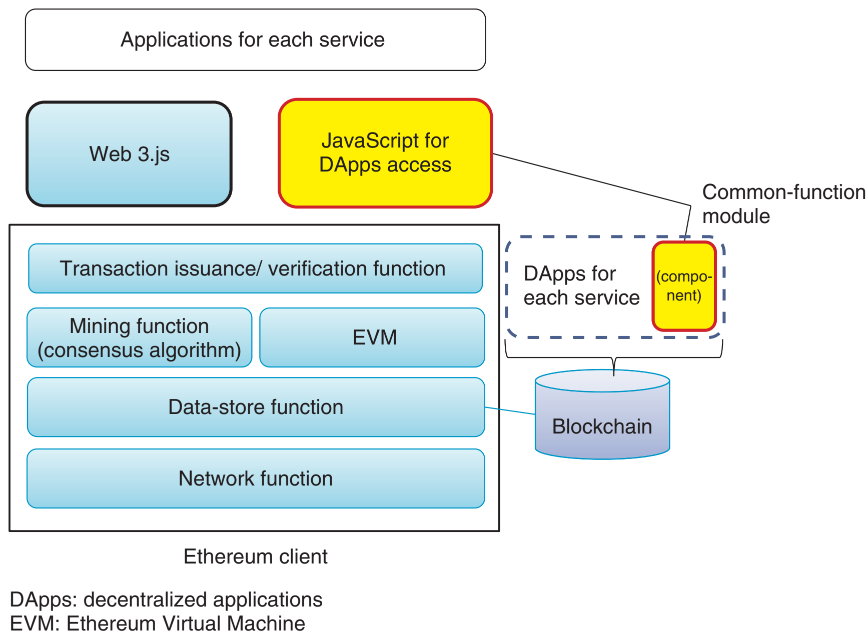


Fig. 2. Proposed module for managing authority and concealing data used in developing blockchain applications.

of the destination and register it in a blockchain. This makes it possible to disclose the common key at the destination only. We are currently investigating how to manage access to externally managed content itself.

Traceability of data is generally achieved by recording data in blockchains and referencing that data by means of a specific ID like a hash value. However, for content, it is necessary to ensure compatibility with derivative works. As for such derivative works, by incorporating a mechanism for confirming the existence (or non-existence) of permission from the original creator of content in distributed applications, we can ascertain the legitimacy of derivative works in terms of rights. Moreover, by confirming features as timestamps of blockchains and details of the above-mentioned registered data, it is possible to manage the originality of content.

4. Support for application development

We are implementing the functions of the above-mentioned technical items in our development of blockchain applications. In the meantime, when efficiency of development on the application layer (i.e., within the code of smart contracts) is considered, it becomes clear that shared functions should be modularized. For example, in use cases other than management of digital content, functions that conceal information in blockchains (such as methods for exchanging encryption keys for content) must also be widely shared. Transparency, correctness, and traceability of transactions are often mentioned as advantages of blockchains. However, in actual business situations, data that cannot be disclosed to other companies are present among customer information and transaction data between businesses. Consequently, it is imperative to protect such highly confidential data.

To satisfy this requirement, it is effective to incorporate functions for concealing information and restricting the right of access. At present, however, methods for implementing such functions in contracts are entrusted to individual developers. It is possible that depending on how the developer writes the code, vulnerable code will be written, and/or dispersibility by blockchains will be weak when the content management is implemented by a central server. In particular, it is difficult for developers who are unused to developing blockchains to appropriately design and implement contracts. We aim to solve this problem while applying our accumulated experience in utilizing blockchains for content management and are therefore proposing and implementing a shared module for creating a framework for managing authority within code, concealing certain stored data variables, and other functions (**Fig. 2**).

The proposed module for managing authority and concealing data is compatible with Ethereum (a type of blockchain platform). With Ethereum, the code of contracts is written in a proprietary language called Solidity. Accordingly, the proposed module is composed of Solidity code (by which the methods for authority management and information concealment are created as templates) and JavaScript code (which provides an interface compatible with that template). By adding the required logic based on the template of the proposed module, the developer can easily develop contracts for handling authority management and information concealment. The functions provided by that template for managing authority management and concealing information are explained in the following.

The template for providing the authority-management function uses a manager contract (for managing access to contracts). The developer registers function calls to which access should be restricted and addresses of contracts with stored data in the manager contract. When the authority-management template is used, the user references individual contracts invariably through the manager contract; as a result, it is possible to centrally restrict access to individual contracts.

To create the template for concealing information, a method for encrypting and storing the values of contracts (that are stored in blockchains) is utilized, and a method for securely managing keys for deciphering those encrypted values is provided. The encryption on the contracts is executed using a common-key method, and the common key is encrypted and stored by means of the public key of each user holding reference authority. The user with authority for decryption executes a function for acquiring a common key encrypted on a blockchain via a JavaScript interface compatible with the template and deciphers the common key by using his/her own private key. In that way, the concealed information can be displayed and viewed by that user.

Modularizing the above-described common functions, regardless of external systems, makes it possible to improve the efficiency of developing applications for disclosing contract data stored in blockchains to the specified user only and applications for the managing authority to overwrite certain data. We believe that creating sufficient functions for modularization and library creation as common functions will contribute to improving the efficiency of development and the quality of applications utilizing blockchains.

5. Concluding remarks

This article described NTT's initiatives concerning blockchains from the viewpoint of the application layer. In particular, factors and items that must be investigated when designing applications using blockchains were explained as use cases exploiting the benefits of blockchains. Implementation methods for managing content were introduced as concrete examples. Additionally, issues concerning the modularization of basic functions and actual implementation of those functions to improve the efficiency of application development were described.

References

[1] S. Nakamoto, "Bitcoin: A Peer-to-peer Electronic Cash System," 2008.
https://bitcoin.org/bitcoin.pdf

[2] S. Underwood, "Blockchain Beyond Bitcoin," Communications of the ACM, Vol. 59, No. 11, pp. 15–17, 2016.

[3] Ethereum Community, "A Next-generation Smart Contract and Decentralized Application Platform," 2014.
https://github.com/ethereum/wiki/wiki/White-Paper

[4] A. Akutsu, K. Hidaka, M. Inoue, N. Ito, T. Yamaguchi, S. Fujimura, and A. Nakadaira, "Delivering Technologies for Services that Deliver the Excitement of Games Worldwide," NTT Technical Review, Vol. 13, No. 7, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507fa2.html

[5] S. Fujimura, H. Watanabe, A. Nakadaira, T. Yamada, A. Akutsu, and J. Kishigami, "BRIGHT: A Concept for a Decentralized Rights Management System Based on Blockchain," Proc. of the 2015 IEEE 5th International Conference on Consumer Electronics - Berlin, pp. 345–346, Berlin, Germany, Sept. 2015.

[6] H. Watanabe, S. Fujimura, A. Nakadaira, Y. Miyazaki, A. Akutsu, and J. Kishigami, "Blockchain Contract: Securing a Blockchain Applied to Smart Contracts," Proc. of the 2016 IEEE International Conference on Consumer Electronics, Las Vegas, NV, USA, Jan. 2016.

Trademark notes

Atsushi Nakadaira

Senior Research Engineer, NTT Service Evolution Laboratories.

He received a B.S. in physics from Kyoto University in 1992, an M.S. in physics from the University of Tokyo in 1994, and a Ph.D. in applied physics from the University of Tokyo in 2001. He joined NTT in 1994 and has been studying III-nitride semiconductors, display devices of polymer distributed liquid crystals, interactive systems of three-dimensional displays, and visual communication systems. He has been involved in research on content management systems leveraging blockchain technology since 2014. He is a member of the Japan Society of Applied Physics and the Institute of Image Information and Television Engineers.

Shigenori Ohashi

Research Engineer, NTT Service Evolution Laboratories.

He received an M.S. in astrophysics from Tohoku University in 2009 and joined NTT the same year. He has been engaged in research and development (R&D) of a virtual co-presence system, metadata management system, and multi-device cooperation system. He has been conducting research on content management systems leveraging blockchain technology since late 2015.

Hiroki Watanabe

Research Engineer, NTT Service Evolution Laboratories.

He received an M.S. in engineering from Waseda University, Tokyo, in 2011 and joined NTT the same year. Since then, he has been researching human interfaces of interactive television (TV), a distribution system for web applications, and content management systems leveraging blockchain technology. His research interest is smart contract systems using blockchain technology.

Shigeru Fujimura

Senior Research Engineer, NTT Service Evolution Laboratories.

He received an M.S. in information science and technology from the University of Tokyo in 2005 and joined NTT the same year. Since then, he has been engaged in research on web mining and web engineering, especially on effective methods of implementing web applications. He is a member of the W3C (World Wide Web Consortium) Web and TV Interest Group.

Satoshi Sakuma

Senior Research Engineer, Supervisor, NTT Service Evolution Laboratories.

He received an M.S. in instrumentation engineering from Keio University, Kanagawa, in 1995 and joined NTT the same year. Since then, he has been engaged in R&D of medical image processing and object recognition. He has also been conducting research on service visualization using NTT's advanced technology since 2014. He received the SPIE Medical Image Processing 1998 Best Poster Award.

Shingo Kinoshita

Executive Research Engineer, Supervisor, and Project Director, NTT Service Evolution Laboratories.

He received a B.E. from Osaka University in 1991 and an M.Sc. with Distinction in technology management from University College London, UK, in 2007. Since joining NTT laboratories in 1991, he has been engaged in R&D of distributed computing systems, security, big data computing, and machine learning. He was a senior manager of the R&D planning section of the NTT holding company from 2012 to 2015, where he established and operated NTT Innovation Institute, Inc. in North America and managed R&D alliance and venture investments. He is presently in charge of the overall direction of various R&D experimental activities toward 2020 including assistance services for foreigners and entertainment services such as kabuki and SXSW. He received the 2005 IPSJ R&D Award from the Information Processing Society of Japan, the 2003 CSS (Computer Security Symposium) Best Paper Award, and the 1998 DICOMO (Multimedia, Distributed, Cooperative, and Mobile) Symposium Best Presentation Award.

Virtual Machine Management Technology for Operating Packet Switching System in a Virtualization Environment

Makoto Hamada, Eriko Iwasa, and Keiko Kuriu

Abstract

NTT Network Service Systems Laboratories is developing infrastructure technology for operating virtualized packet switching system service functions with the aim of realizing a network control infrastructure providing safe and secure services. In this article, we introduce virtual machine management technology for sharing hardware resources in a virtualization environment and onboarding and operating packet switching system service functions in conjunction with OpenStack.

Keywords: virtual machine management, packet switching, network functions virtualization

1. MAGONIA rollout

The NetroSphere concept [1] was announced by NTT in February 2015 as a new research and development concept aimed at revolutionizing the way to form a carrier network infrastructure. As part of this initiative, NTT Network Service Systems Laboratories is researching and developing a new server architecture called MAGONIA to enable the short-term creation and deployment of services and to achieve drastic reductions in development and operating costs.

MAGONIA provides a distributed processing base (DPB) [2, 3], Soft Patch Panel (SPP) technology [4], and virtual machine (VM) management technology as components that can be used by individual applications. In this article, we introduce MAGONIA VM management technology.

Studies on network functions virtualization (NFV) are moving forward in the European Telecommunications Standards Institute Industry Specification Group for NFV (ETSI NFV ISG). NFV is an initiative to run VMs on general-purpose servers and to provide diverse services in a flexible and prompt manner by virtualizing (converting to software) network functions traditionally provided by dedicated hardware appliances. Virtualization makes it possible to run more than one virtually created machine on a single piece of hardware, which means that VMs can be dynamically created as needed, with each one sized most appropriately for the conditions at hand.

The ETSI NFV architecture is divided into three main components, as shown in **Fig. 1**: network functions virtualization infrastructure (NFVI), virtual network functions (VNF), and management and orchestration (MANO). The NFVI area consists of physical resources and virtualization functions for running VNF. It achieves a function called a hypervisor for generating multiple virtual servers on a physical server. The kernel-based VM (KVM) is a typical hypervisor. The VNF area, meanwhile, consists of network functions implemented as software. Finally, the MANO area, whose central role is operation automation, consists of functional blocks called the virtualized infrastructure manager (VIM), VNF manager (VNFM), and NFV orchestrator, each with separate roles.

The VIM manages NFVI and allocates VMs, the

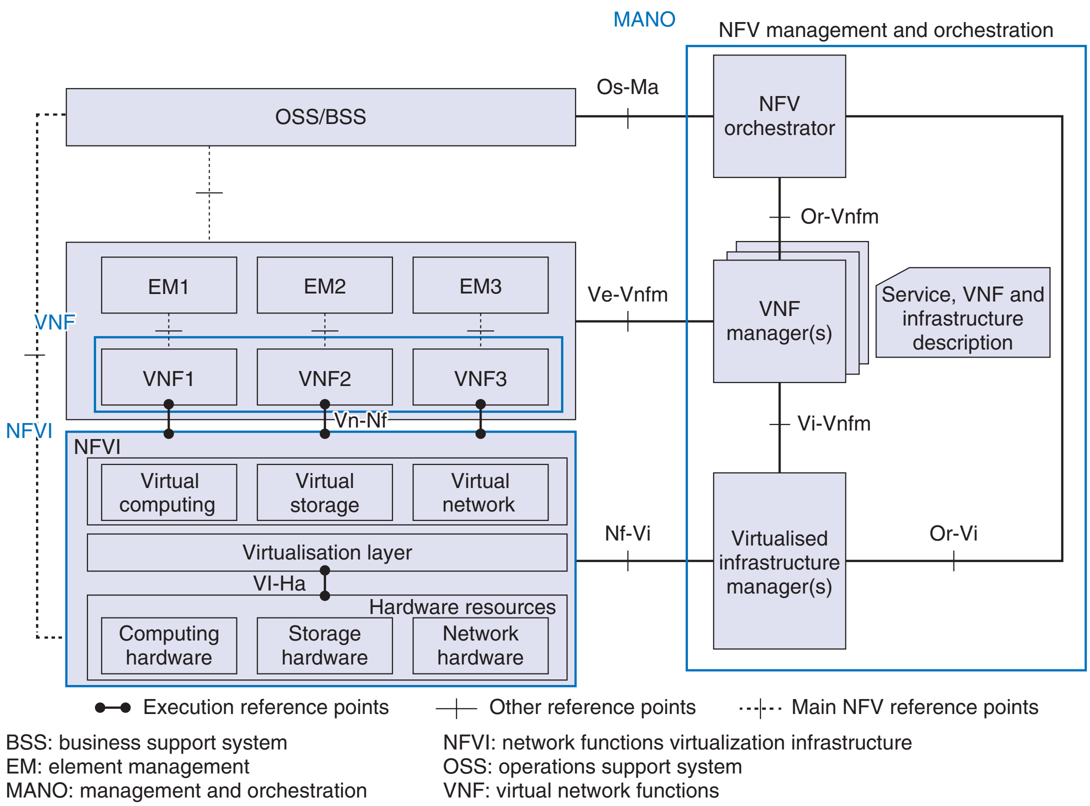


Fig. 1. ETSI NFV architecture.

VNFM manages VNF, and the NFV orchestrator manages, controls, and optimizes network services and resources in an integrated manner. While OpenStack has become the de facto standard for achieving a VIM, NFV products used throughout the world consist of original vendor developments based on OpenStack. Various telecommunications carriers are focusing on the application of NFV architecture to large-scale communications networks, but it is important that issues surrounding the development and implementation of packet switching service functions are solved.

2. Issues in virtualization of packet switching system service functions

The role of network packet switching system service functions such as firewall and network address translation (NAT) functions is to check and process the contents of all received packets. It is therefore necessary to process a large volume of packets with as short a delay as possible. Such packet switching system service functions have traditionally been implemented using dedicated equipment to satisfy this strict performance requirement. In MAGONIA, these functions are converted into software (packet switching software) and operated on a common platform based on virtualization technology with the aim of reducing capital expenditure and operating expenses and invigorating service development.

However, a number of issues arise in virtualizing packet switching system service functions. Since virtualization technology achieves multiple VMs on a single piece of hardware, these VMs can interfere with each other and cause processing delay and unstable throughput. For a server system supporting web services, which is also targeted by virtualization technology, latency requirements are not strict, and it has been possible to configure a system that increases the number of hardware units whenever processing power becomes insufficient, even if throughput is unstable.

Packet switching system service functions, however, must process a large volume of packets with no allowance given for processing delay. In addition, many existing communications networks have been

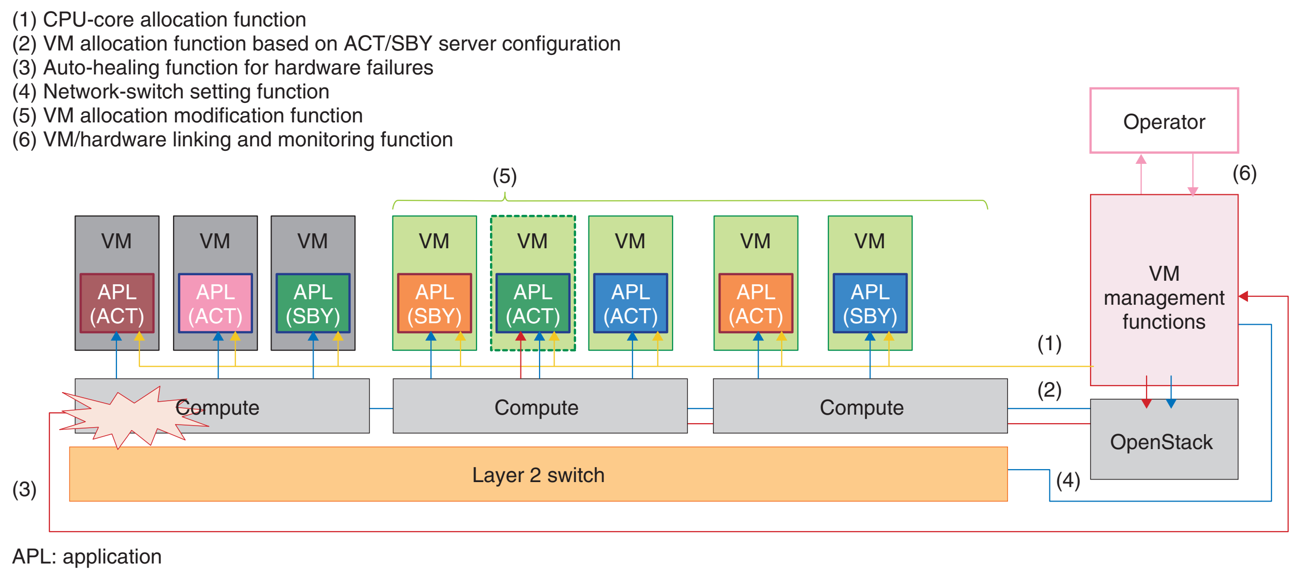


Fig. 2. Functions achieved by VM management technology.

constructed with a specific type of architecture that sets beforehand the number of users that can be processed by each piece of equipment, so for this and the above reason, performance cannot be maintained by simply increasing the number of hardware units in the case of web servers. Thus, when implementing packet switching system software using NFV architecture, the performance of virtualized service functions must be guaranteed assuming that type of architecture.

In addition, existing communications networks consist of active and standby (ACT/SBY) servers and employ a failover mechanism in which a SBY server takes over the processing of an ACT server that fails. Under a virtualization environment, this means that VM allocation control must take into account a system having a redundant-configuration mechanism. The VM management technology in MAGONIA achieves functions for solving those issues that arise in applying network services to a virtualization environment.

3. Functions achieved by VM management technology

The functions achieved by VM management technology in MAGONIA are shown in **Fig. 2**. The central processing unit (CPU)-core allocation function (1), which is one of the KVM functions, allocates VM processing to CPU cores. This function achieves low processing delay and stable throughput by exclusively allocating processes requiring high-speed processing to CPU cores. Next, the VM allocation function based on an ACT/SBY server configuration (2) can allocate VMs to optimal hardware in a system having a redundant configuration. In addition, the auto-healing function for hardware failures (3), the network-switch setting function (4), the VM allocation modification function (5), and the VM/hardware linking and monitoring function (6) support failover processing at the time of a hardware failure or fault occurrence and lifecycle management tasks.

4. CPU-core allocation function in MAGONIA VM management technology

The mechanism of the CPU-core allocation function is shown in **Fig. 3**. This function provided by MAGONIA VM management technology has three separate functions.

The first function allocates the processing of one virtual CPU within a VM to one physical CPU. In this way, it exclusively allocates a physical CPU core to a virtual CPU executing high-load processing, thereby maintaining the processing performance of that virtual CPU.

The second function exclusively allocates a group of virtual CPUs to a single physical CPU core. Applying this function to multiple virtual CPUs performing processing that consists of a light load on the machine can achieve effective use of physical CPU resources

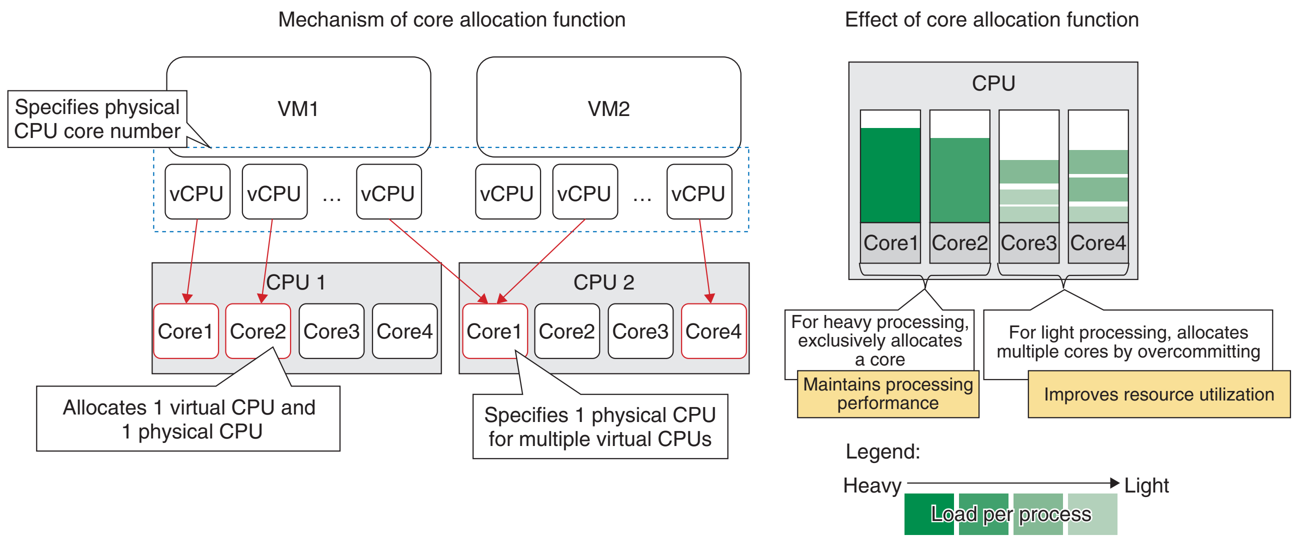


Fig. 3. CPU-core allocation function.

and ensure adequate processing time.

The third function specifies a certain core number in the physical CPU allocated to a virtual CPU. Performance can differ between CPU cores according to the network interface card or socket connected to each core, so specifying a core number enables core allocation that takes performance requirements into account. This function therefore makes it possible to maintain the processing performance of an allocated virtual CPU or avoid a drop in performance due to interruption of operating system processing.

To achieve this function, it must be set in KVM instead of VIM. However, in OpenStack, the VIM de facto standard, there is no CPU-core allocation function with a granularity that can be set by KVM. For this reason, VM management functions use OpenStack as the VIM and implement a function for allocating the CPU cores needed for operating packet switching system service functions.

5. Effect of VM management technology

The results of increasing the number of VMs generated by VM management technology and measuring the number of packets that can be normally processed while supplying large volumes of packets to those VMs are shown in **Fig. 4**. The plot points show that throughput performance increases linearly as the number of VMs becomes larger when applying VM management technology. It can therefore be seen that application of this CPU-core allocation function generates VMs that ensure required performance.

This VM management technology can also allocate CPU cores at a quantity needed for an ACT/SBY configuration as well as for a VM used for auto-healing at the time of a hardware failure to ensure stable provision of services immediately after system recovery.

6. Future development

The VM management technology of MAGONIA is expected to be useful for virtualizing functions having severe latency requirements and requiring stable throughput as in the case of packet switching system service functions. We aim to develop these functions into common functions for supporting the future server infrastructure; therefore, our plan is to expand associated applications and engage in proposal activities with OpenStack and other open source software communities.

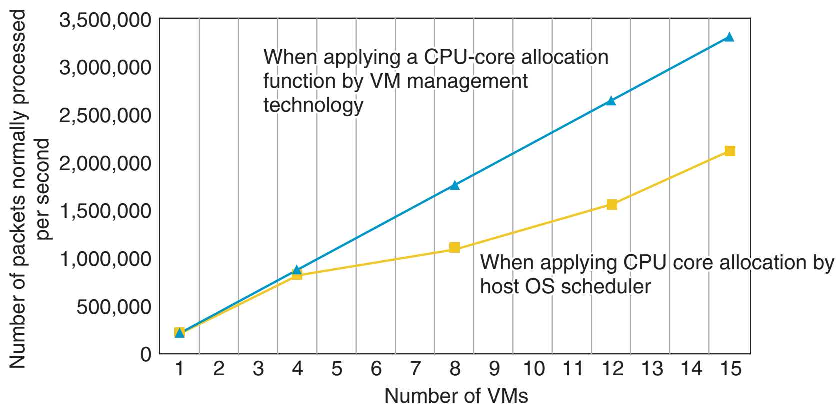


Fig. 4. Sample throughput by VM management technology.

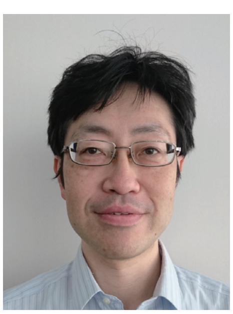
References

[1] T. Ito, "Accelerating the Demonstration and Adoption of NetroSphere," NTT Technical Review, Vol. 14, No. 10, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201610fa1.html

[2] H. Kobayashi, T. Kitano, M. Okamoto, and T. Fukumoto, "MAGONIA (DPB: Distributed Processing Base) Applied to a Traffic Congestion Prediction and Signal Control System," NTT Technical Review, Vol. 14, No. 10, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201610fa7.html

[3] NTT, "Sophisticated Network-node Integration Architecture," NTT R&D Forum, E-1, Feb. 2012.
https://www.ntt-review.jp/library/ntttechnical.php?contents=ti1202rdfpa25e.pdf

[4] Soft Patch Panel - DPDK resources management framework,
http://www.dpdk.org/browse/apps/spp/



Makoto Hamada

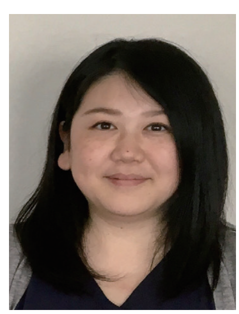
Research Engineer, Server Network Innovation Project, NTT Network Service Systems Laboratories.

He received his B.S. and M.S. in mechanical engineering from Tokyo Institute of Technology in 1991 and 1993. He joined NTT Telecommunication Networks Laboratory in 1993, where he worked on the research and development of communication node systems, including Next Generation Network (NGN) session control servers. He is currently studying distributed systems and network functions virtualization. He is a member of the Institute of Electronics, Information and Communication Engineers.

Eriko Iwasa

Research Engineer, Server Network Innovation Project, NTT Network Service Systems Laboratories.

She received an M.E. in electrical engineering from Waseda University, Tokyo, in 2009 and joined NTT the same year, where she was engaged in researching distributed computing technology. She is currently studying network functions virtualization.



Keiko Kuriu

Research Engineer, Server Network Innovation Project, NTT Network Service Systems Laboratories.

She received an M.S. in information and computer sciences from Nara Women's University in 2012 and joined NTT the same year. She began her career promoting distributed computing technology. She is currently studying network functions virtualization.

Report on ITU-T Study Group 15 Meeting

Noriyuki Araki, Makoto Murakami, and Jun-ichi Kani

Abstract

ITU-T (International Telecommunication Union - Telecommunication Standardization Sector) Study Group (SG) 15 was reorganized for the study period from 2017 to 2020. The new structure of SG15, which includes 3 Working Parties (WPs) and 19 Questions, has been approved. WP chairmen, vice-chairmen, and rapporteurs have been appointed. The results of the first SG15 meeting for this study period and the activity plans for the Questions are reported in this article.

Keywords: ITU-T, SG15, optical technology

1. Introduction

The International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) is an organization dedicated to the international standardization of telecommunication technologies. ITU-T updates its structure every four years, which is considered to be the basic study period. This article describes the reorganization of Study Group (SG) 15 for the study period from 2017 to 2020 and the results of the group's first meeting in this period, which was held in Geneva, Switzerland, in June 2017.

2. Organization

The chairmen and vice-chairmen of all SGs for the study period from 2017 to 2020 were elected at the ITU World Telecommunication Standardization Assembly held in Tunisia in October 2016. One chairman and ten vice-chairmen of SG15 were elected (**Table 1**). The first SG15 meeting of the new study period was held in Geneva from June 19–30, 2017. The new structure of SG15, which consists of 3 Working Parties (WPs) and 19 Questions (Qs), was approved.

3. WPs and Questions

SG15 is one of the largest and most active SGs in ITU-T, and it covers a very wide study area that includes optical/metallic transmission media, optical physical infrastructure and transmission systems in access and core networks, and network logical layers such as packet transmission and synchronization. Formal approval was given for this SG to conduct standardization work with three WPs, just as in the previous study period. The WPs will focus on the transport aspects of access, home, and smart grid networks (WP1), optical technologies and physical infrastructure (WP2), and transport network characteristics (WP3). Although some Question titles were changed, all the original Questions from the previous study period remained unchanged except for the transfer of a Question related to home network technology from SG9 (CATV: cable television). The WP structures, WP chairmen and vice-chairmen, and the Questions and rapporteurs in SG15 are listed in **Table 2**.

Delegates from NTT were appointed WP2 chairman, rapporteurs of Q3, Q5, and Q17, and Q2 associate rapporteur. This indicates that NTT members will lead many of the main discussions in SG15.

4. Summary of first SG15 meeting

The first SG15 meeting was attended by 272 delegates representing 35 countries. There were more participants and countries represented than at the last

Table 1. SG15 chairman and vice-chairmen for study period 2017–2020.

Title	Name	Country	Term
Chairman	Steve Trowbridge	USA	Reappointment
Vice-chairmen	Noriyuki Araki	Japan	Reappointment
	Dan Li	China	Reappointment
	Jeong-Dong Ryoo	Republic of Korea	Reappointment
	Fahad Abdullah Al-Fallaj	Saudi Arabia	Reappointment
	Khaled Al-Azemi	Kuwait	Newly appointed
	Hubert Mariotte	France	Newly appointed
	Cyrille Vivien Vezongada	Central African Republic	Newly appointed
	Glenn Parsons	Canada	Newly appointed
	Edoardo Cottino	Italy	Newly appointed
	John Messenger	United Kingdom	Newly appointed

SG15 meeting. Thirty delegates from Japan attended the meeting, which was the third highest number of participants after China and the United States. The meeting received 389 contributions and 482 Temporary Documents. The Japanese delegates submitted 26 contributions, the same as at the last meeting. SG15 approved 2 Recommendations, consented to 52 texts for draft Recommendations, Amendments, and Corrigenda, and agreed on 7 texts, including Supplements and Technical Reports. The main discussion results of the meeting, categorized by WP, are described below.

5. WP1: Transport aspects of access, home, and smart grid networks

WP1 studies optical and metallic access systems, home networks, and smart grid networks.

Q1 addresses the production and updating of standardization outlines of access and home networks and standardized work plans through information exchanges between ITU-T SGs and other standards development organizations (SDOs).

Q2 examines optical access systems such as passive optical networks (PONs). In the previous study period, G.989 series Recommendations for 40-Gbit/s next-generation optical access systems (NG-PON2: Next-Generation Passive Optical Network Stage 2) were completed. At this meeting, amendments to the new Recommendation G.9807.2, which deals with 10-Gbit/s symmetric PON (XGS-PON: 10-Gigabit-capable symmetric passive optical network) and the existing Recommendation related to NG-PON2, received consent; their specifications have been expanded in accordance with carrier requirements. Progress was also made on the new Recommendation G.RoF for radio over fiber technologies with a view to receiving consent after the next meeting. Q2 members decided to start work on a supplement of 5G (fifth-generation mobile communications systems) fronthaul requirements in a PON context.

Q4 deals with metallic access systems. In the previous study period, G.fast Recommendations (G.9700 series) were produced that dealt with ultrahigh-speed metallic access systems with 1-Gbit/s transmission capacity. At this meeting, progress was made on the revision of the relevant Recommendations, which included the approval of an amendment to Recommendation G.9700 with a frequency profile specification of the 212-MHz band. Q4 members also agreed to begin working on Recommendation G.mgfast (multi-Gigabit fast access to subscriber terminals) for a higher bit rate as the next-generation transceiver specification.

Q15 deals with narrowband power line communication for smart grid networks. The revised Recommendation G.9901, which stipulates the output voltage level, received consent at this meeting.

Q18 examines Recommendations for home networking. At this meeting, Q18 members produced amendments to G.hn series Recommendations, which include an overhaul of network authentication protocols and provisions for domain master selection. It was also agreed to start a new project on G.hn2.0 to support high-speed home networking.

Q19 addresses CATV home networks. However, SG15 agreed to merge Q19 with Q18, which also studies home networking, because there was no participating delegate from SG9 and the Q19 rapporteur position is still vacant. SG15 sent a Liaison Statement

Table 2. WP structure and list of Questions and rapporteurs.

Question	Title	Rapporteur
WP1: Transport aspects of access, home, and smart grid networks (Chairman: Tom Starr, AT&T, USA) (Vice-chairman: Hubert Mariotte, Orange, France)		
Q1	Coordination of access and home network transport standards	R) J-M Fromenteau, Corning, USA AR) Dekun Liu, Huawei, China
Q2	Optical systems for fibre access networks	R) Frank Effenberger, Futurewei Technologies, USA AR) Jun-ichi Kani, NTT, Japan
Q4	Broadband access over metallic conductors	R) Frank Van Der Putten, Nokia, Finland AR) Les Brown, Huawei, China AR) Miguel Peeters, Broadcom, USA
Q15	Communications for smart grid	R) Stefano Galli, Huawei, China AR) Paolo Treffiletti, STMicroelectronics, Italy
Q18	Broadband in-premises networking	R) Les Brown, Huawei, China AR) Marcos Martinez, Maxlinear, USA
Q19	Requirements for advanced service capabilities over broadband cable home networks	Vacant
WP2: Optical technologies and physical infrastructures (Chairman: Noriyuki Araki, NTT, Japan) (Vice-chairman: Pete Anslow, Ciena, Canada)		
Q5	Characteristics and test methods of optical fibres and cables	R) Kazuhide Nakajima, NTT, Japan AR) David Mazzarese, OFS, USA
Q6	Characteristics of optical systems for terrestrial transport networks	R) Peter Stassar, Huawei, China AR) Pete Anslow, Ciena, Canada
Q7	Characteristics of optical components and subsystems	R) Bernd Teichmann, Nokia, Finland
Q8	Characteristics of optical fibre submarine cable systems	R) Omar Ait Sab, Nokia, Finland
Q16	Optical physical infrastructures	R) Edoardo Cottino, SIRTI, Italy AR) Osman Gebizlioglu, Huawei, China
Q17	Maintenance and operation of optical fibre cable networks	R) Kunihiro Toge, NTT, Japan AR) Xiong Zhuang, MIIT, China
WP3: Transport network characteristics (Chairman: Malcolm Betts, ZTE, China) (Vice-chairman: Glenn Parsons, Ciena, Canada)		
Q3	Coordination of optical transport network standards	R) Naotaka Morita, NTT, Japan
Q9	Transport network protection/restoration	R) Tom Huber, Coriant, Germany
Q10	Interfaces, interworking, OAM and equipment specifications for packet-based transport networks	R) Jessy Rouyer, Nokia, Finland
Q11	Signal structures, interfaces, equipment functions, and interworking for optical transport networks	R) Steve Gorshe, Microsemi, USA
Q12	Transport network architectures	R) Stephen Shew, Ciena, Canada
Q13	Network synchronization and time distribution performance	R) Stefano Ruffini, Ericsson, Sweden AR) Silvana Rodrigues, IDT, Canada
Q14	Management and control of transport systems and equipment	R) Hing-Kam Lam, Fiberhome, China AR) Scott Mansfield, Ericsson, Canada

R) Rapporteur, AR) Associate Rapporteur

to seek endorsement at the next Telecommunication Standardization Advisory Group (TSAG) meeting.

6. WP2: Optical technologies and physical infrastructure

WP2 studies physical interfaces, transmission characteristics, technologies related to construction, and the maintenance and operation of outside plants in optical transport networks. At this meeting, the Technical Report of the guide to L-series Recommendations was revised with a new numbering system and the addition of new Recommendations.

Q5 examines the characteristics and test methods of optical fiber and cables. In the previous study period, members made an effort to accommodate requests regarding transmission systems and therefore revised Recommendations G.652 and G.657—respectively related to conventional single-mode fiber and bending-loss insensitive single-mode fiber—in terms of wavelength dispersion characteristics. Q5 members also agreed to the expansion of the applicable areas of G.657 fiber and revised G.654 by adding a new fiber category for terrestrial coherent transmission systems. At this meeting, Recommendation G.650.3 dealing with test methods for installed optical fiber cable links was also revised.

Q6 examines the characteristics of optical systems for terrestrial transport networks. After discussions on the revision of Recommendation G.698.2 related to dense wavelength division multiplex (DWDM) applications, it was agreed that the target consent date for the revised G.698.2, which includes additional 100-Gbit/s applications, will be October 2018. Existing Recommendations related to Ethernet standards and DWDM systems intended for metro (metropolitan) applications were also discussed.

Q7 focuses on Recommendations related to optical subsystems and optical components. At this meeting, a new Recommendation L.404 describing a field mountable single-mode optical fiber connector received consent. This type of connector is widely used in NTT's facilities as a FAS (field assembly small) connector and for field assembly wavelength selective terminations.

Q8 focuses on optical fiber submarine cable systems. G suppl. 41, which provides a guideline for the design of optical fiber submarine cable systems, was discussed at the meeting.

Q16 covers the overall optical physical infrastructure including optical fiber cables and examines Recommendations related to the characteristics of optical fiber cable systems such as optical performance and mechanical and environmental test conditions, for the purpose of promoting FTTH (fiber-to-the-home). At this meeting, Q16 members consented to new Recommendations L.110, optical fiber cable for direct surface applications, and L.202, outdoor optical cross-connect cabinet. The members also agreed to start work on standards for optical fiber cables for in-home applications.

Q17 deals with the maintenance and operation of optical fiber cable networks. At this meeting, L suppl. 35 was agreed to as a framework of disaster management for network resilience and recovery. Q17 members also discussed new Recommendations related to a water detection module for underground closures and live fiber detection technology for PONs.

7. WP3: Transport network characteristics

WP3 mainly studies the logical layers of transport networks and focuses on the seven Questions described below. At this meeting, members discussed wide-ranging technologies such as packet based transport networks (e.g., Ethernet and Multi Protocol Label Switching-Transport Profile (MPLS-TP)), network architecture such as beyond 100-Gbit/s optical transport network (OTN) and transport software-defined networking (SDN), and network synchronization.

Q3 members are working on coordinating effective optical transport network standards. They updated the optical transport networks and technologies standardization work plan and sent Liaison Statements to the relevant SDOs for collaboration and information exchange.

Q9 examines protection and restoration technologies for transport networks. Members discuss conventional protection and restoration issues and issues related to Ethernet, MPLS-TP, and OTN. Consent was given at this meeting to two new Recommendations: G.873.3 for OTN mesh protection and G.8132 for the ring protection of MPLS-TP.

Q10 examines service interfaces, OAM (operations and administration and maintenance) mechanisms, and equipment requirements for packet transport technologies of Ethernet and MPLS-TP. Consent was given to some amendments to the existing Recommendation related to the MPLS-TP interface.

Q11 mainly addresses signal structures, interfaces, equipment functions, and interworking for OTN. Q11 members made progress in their discussions on over-100-Gbit/s OTN and refined the document structure of Recommendation G.709.1 to add 200/400-Gbit/s applications. Also, in a joint meeting with other relevant Question teams, members discussed the requirements of transport networks for International Mobile Telecommunication (IMT)-2020/5G because the evolution of the transport network to support IMT-2020/5G is an important new topic. The members of the joint meeting shared a common view in SG15 in regard to front/backhaul network classification, application scenarios, and the characteristics of transport networks for supporting IMT 2020/5G scenarios. These scenarios are based on discussions held by the 3rd Generation Partnership Project (3GPP). To avoid misleading the discussion, it was agreed to send a Liaison Statement to 3GPP to confirm whether the SG15 common view is correct. WP3 agreed to start work on a new Technical Report on the transport network to support IMT-2020/5G.

Q12 addresses the transport network architecture for OTN and the implementation of SDN and network functions virtualization in an actual transport network. Q12 members proceeded with work on creating a Recommendation for transport SDN control architecture as the result of discussions on transport network architecture for IMT-2020/5G, SDN, and ASON (automatically switched optical network) architecture at this meeting.

Q13 examines the standardization of frequency synchronization for transport networks and time and frequency synchronization for packet networks. To accommodate the increase in network traffic caused by IPs (Internet protocols) such as MPLS-TP and the requirements for IMT-2020/5G, the members of Q13 made progress on Recommendations for higher-precision time and frequency synchronization technologies.

Q14 deals with Recommendations for network management and control technologies. Management requirements for general transport network elements and management requirements and an information model for specific technologies such as OTN, Ethernet, and MPLS-TP were discussed. Recommendation G.874, management aspects of OTN elements, and some amendments to the relevant Recommendations were revised at this meeting.

8. Future activities

The second SG15 plenary meeting in the 2017–2020 study period will be held in Geneva from January 29 to February 9, 2018. To allow sufficient discussion, SG15 has planned many interim meetings and e-meetings before that meeting. Also, an ITU-T SG15, IEEE (Institute of Electrical and Electronics Engineers) 802.3, and 802.1 joint interim meeting and joint workshop are planned for January 22–26 and January 27, 2018, respectively, just prior to the next SG15 meeting.

Noriyuki Araki

Senior Research Engineer, Access Media Project, NTT Access Network Service Systems Laboratories.

He received a B.E. and M.E. in electrical and electronic engineering from Sophia University, Tokyo, in 1993 and 1995. He joined NTT Access Network Service Systems Laboratories in 1995, where he researched and developed operation and maintenance systems for optical fiber cable networks. He has been contributing to standardization efforts in ITU-T SG6 since 2006. He was the rapporteur of Question 6 in ITU-T SG6 from 2006 to 2008 and the rapporteur of Question 17 in ITU-T SG15 from 2008 to 2012. He also served as the chairman of the ITU-T Focus Group on Disaster Relief Systems and Network Resilience and Recovery. He has been a vice-chair of ITU-T SG15 since 2013. He also contributes to the activities of IEC (International Electrotechnical Commission) TC86 (Technical Committee 86: Fibre optics). He received the ITU-AJ award from the ITU Association of Japan in 2012. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Makoto Murakami

Senior Research Engineer, NTT Network Service Systems Laboratories.

He received a Ph.D. in electrical engineering from the University of Tokyo in 2009. He initially engaged in research and development (R&D) of long-haul transmission systems using optical amplifiers and coherent modulation/demodulation schemes at the emergence of those technologies. After completing development and deployment of a commercial optically amplified submarine system, he continued R&D of WDM systems to further increase the fiber transmission capacity. From 2001 to 2003, he worked for NTT Communications, where he was involved in the construction and operation of international communication networks mainly in the Asia-Pacific region. Since 2003 he has been an active participant in ITU-T SG15 as head of the Japanese delegation and has also been involved in R&D and standardization of large-capacity optical transport networks. He is currently the chairman of the transport networks and EMC (Electro-Magnetic Compatibility) Working Group in the Telecommunication Technology Committee (TTC) of Japan. He received the Accomplishment Award from the ITU Association of Japan in 2015 and the Distinguished Service Award from TTC in 2015.

Jun-ichi Kani

Senior Research Engineer, Supervisor (Distinguished Researcher), Group Leader, Optical Access System Project, NTT Access Network Service Systems Laboratories.

He received an M.E. and Ph.D. in applied physics from Waseda University, Tokyo, in 1996 and 2005. He joined NTT in 1996. He has been with NTT Access Network Service Systems Laboratories since 2003, where he has been engaged in R&D and standardization of optical communications systems for metro and access applications.

Case Studies of Problems in Digital Cordless Phones

Abstract

This article describes various faults that were occurring with digital cordless phones and the investigations and countermeasures implemented to solve the problems. This is the forty-third article in a series on telecommunication technologies. This contribution is from the EMC Engineering Group, Technical Assistance and Support Center, Maintenance and Service Operations Department, Network Business Headquarters, NTT EAST.

Keywords: digital cordless phone, PHS, radio signal

1. Introduction

The diversification of business formats and means of communication in offices, manufacturing plants, and elsewhere has led to the use of digital cordless phones (1.9-GHz band) such as the Personal Handyphone System (PHS) [1] as a supplement to conventional fixed-line telephone sets. More recently, the introduction of digital cordless phones based on the new Digital Enhanced Cordless Telecommunications (DECT) system (1.9-GHz band) has been progressing in addition to PHS. As a result, digital cordless phones are being used with increasing frequency in diverse environments such as plants and warehouses in addition to offices.

However, this enhanced deployment of digital cordless phones is being accompanied by problems associated with the customer's environment, radio signal level, and interference. In short, the causes of faults in the use of digital cordless phones are becoming more diverse, making early recovery from such faults and the improvement of service quality a matter of urgency.

In this article, we introduce recent case studies of problems in digital cordless phones handled by the Technical Assistance and Support Center.

2. Case study 1: Communication failure inside a store (αNX II business phone)

The first case study involves a communication failure. The problem is first described, and the results of the investigation and the countermeasures implemented are explained.

2.1 Overview and report

A customer using digital cordless phones (PHS: 1.9-GHz band) connected to NTT's αNX II business phone system reported a problem in which calls could not be made or received within the customer's retail store (behind the counter on the first floor: see **Fig. 1**). In this configuration, one cell station (CS) is located near the window on the second floor, and two PHS terminals are in use. Replacing the terminals did not solve the problem.

2.2 Results of measurement and cause of failure

A spectrum analyzer (Keysight Technologies N9020A) was used to measure the signal strength at the center of the store on the second floor and behind the counter on the first floor. The CS is located near the center of the store on the second floor, so we were able to observe signals that had sufficient strength (–40 dBm) at that measurement point (**Fig. 2(a)**).

DECT signals were also observed in addition to PHS signals. At the measurement point behind the counter on the first floor, the signal strength from the

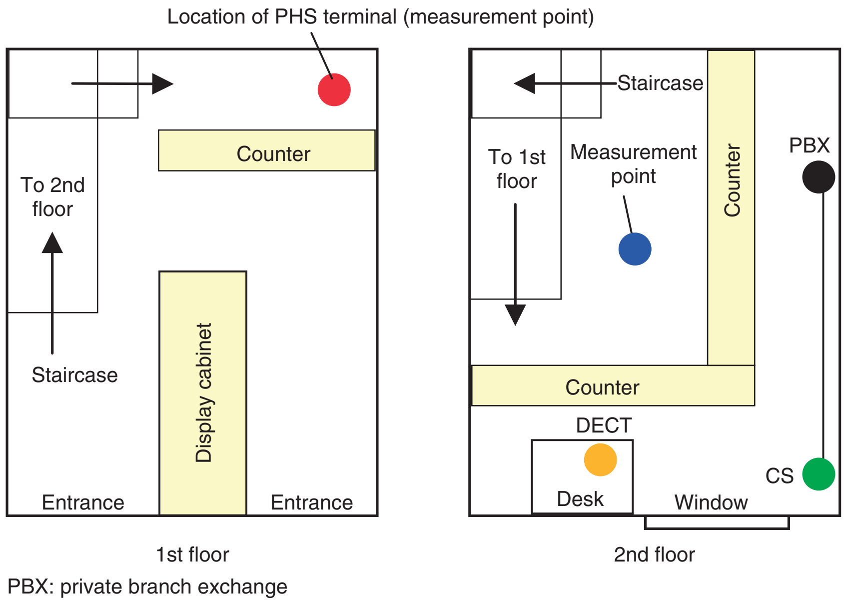


Fig. 1. Equipment configuration (customer's store).

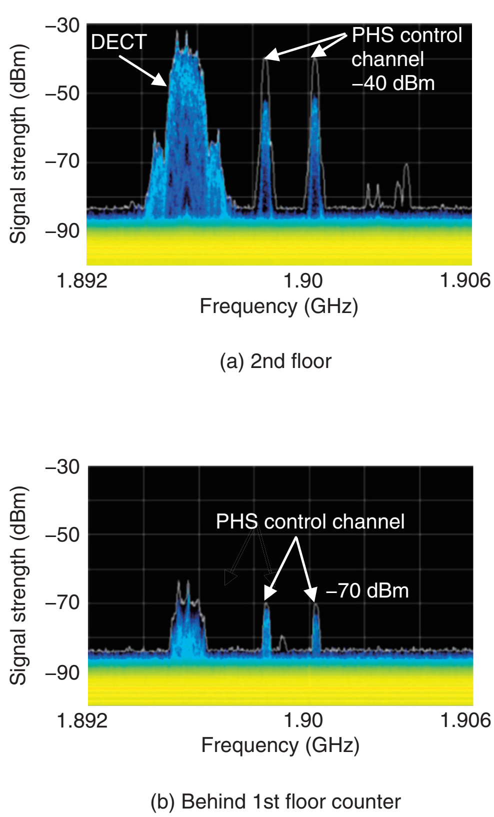


Fig. 2. Results of measuring signal strength.

CS installed on the second floor was found to have dropped to as low as –70 dBm (**Fig. 2(b)**).

The results of the above measurements led us to infer that this problem occurred because the PHS terminals were being used at a location that was too far away from where the CS was installed, resulting in insufficient signal strength at the terminal usage location.

2.3 Countermeasure

We received the customer's consent to add a CS near the location where the PHS terminals were being used (near the first floor counter). We found that stable communications could be achieved with the signal strength of –65 dBm at the terminal usage location.

3. Case study 2: Phone noise in office (PBX EP74)

The second case study involved phone noise. We report here on the problem encountered by the customer, the results of the subsequent investigation, and the recommended countermeasure.

3.1 Overview and report

A customer using a private branch exchange (PBX) (EP74) reported noise and dropped calls when using digital cordless phones (PHS terminals) accommodated

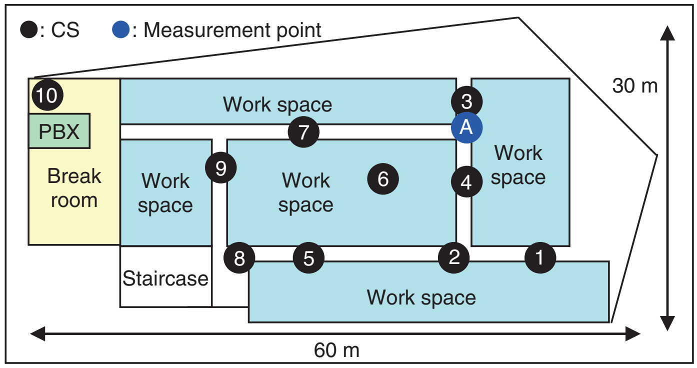


Fig. 3. Floor diagram.

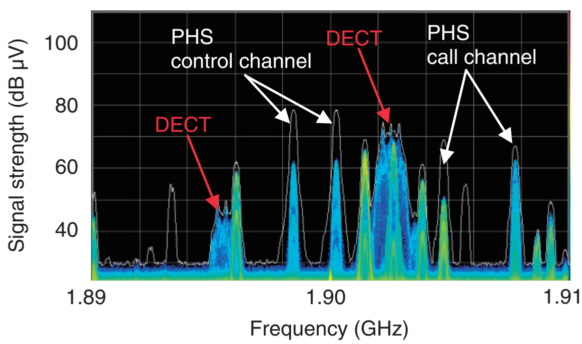


Fig. 4. Radio signal environment (measurement point A).

by the PBX. A total of 10 CSs were installed throughout the floor, and 109 PHS terminals were being used (**Fig. 3**). Replacing terminals failed to solve the problem.

3.2 Results of measurement and cause of failure

We used a spectrum analyzer (Keysight Technologies N9020A) to measure the radio signal environment in the PHS band below CS 3 (see Fig. 3). In this measurement, DECT signals could be observed in addition to PHS control-channel and call-channel signals (**Fig. 4**).

The results revealed that the DECT and PHS signals were interfering with each other. This led us to infer that the cause of noise in the phones was this interference between DECT and PHS signals. Although PHS and DECT have functions for mutually avoiding interference, it can be inferred that a high channel utilization rate may leave no empty channel available for interference avoidance, resulting in the occurrence of interference and dropped calls.

3.3 Countermeasure

To prevent this phone noise from occurring, we recommended to the customer that no DECT terminal be used or alternatively that the DECT terminal be exchanged with a terminal using frequency bands other than the 1.9-GHz band (2.4-GHz digital cordless phones, 5-GHz Wi-Fi devices, etc.).

4. Case study 3: Dropped calls in a food plant (αNX II business phone)

We explain a problem involving dropped calls as the third case study. The results of our investigation into the problem and the countermeasure to rectify it are also reported.

4.1 Overview and report

Dropped calls and line disconnections were occurring in digital cordless phones (PHS: 1.9-GHz band) accommodated by NTT's αNX II-L business phone system. The customer reported that such problems occurred frequently in the processing room of the food plant depicted in **Fig. 5**. Measures such as replacing terminals or adjusting signal power failed to solve the problem.

4.2 Results of measurement and cause of failure

We used a spectrum analyzer (Tektronix RSA6114A) to measure the radio signal environment at the measurement point within the processing room (see Fig. 5). The PHS signal strength was found to be sufficiently strong, and no interfering signals were observed (**Fig. 6**).

The processing room has wall and floor surfaces

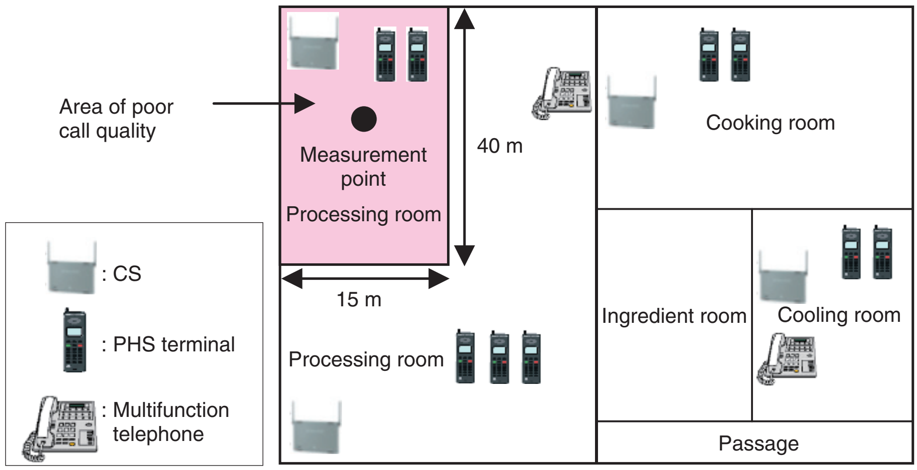


Fig. 5. Equipment configuration (2nd floor of plant); PBX is installed in main distribution frame room (MDF) on third floor.

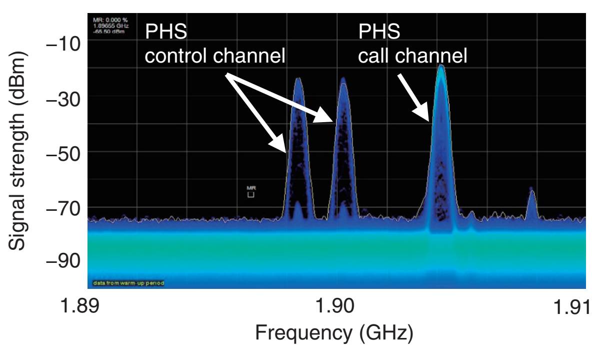


Fig. 6. Results of measuring signal strength.

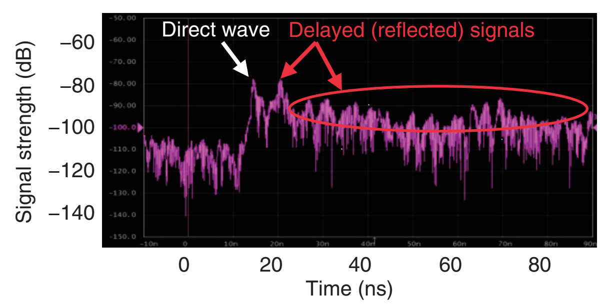


Fig. 7. Results of measuring reflected signals (at time of a dropped call).

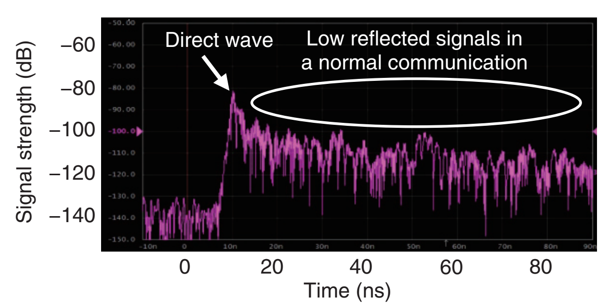


Fig. 8. Results of measuring reflected signals (normal communication).

and machinery made of stainless steel and other types of metal, so we considered that reflected signals might have an effect on call quality. We measured such reflected signals with a network analyzer (Keysight Technologies E5071C) and observed the delay (reflected) signals of nearly the same strength as direct waves and long-delay (reflected) signals of 80 ns or longer (**Fig. 7**). When compared with a location enabling normal communications (**Fig. 8**), the above results revealed the existence of a multipath.

The results led us to infer that dropped calls in the processing room were caused by reflected signals.

4.3 Countermeasure

The customer's processing room with its wall and floor surfaces and machinery made of stainless steel and other metals is conducive to the generation of reflected waves, resulting in a critical environment for radio signals. We explained to the customer that such an environment made the use of digital cordless phones difficult and proposed the use of fixed-line phones. To prevent future problems, we recommended that call tests be performed before introducing a new service to see whether dropped calls or other problems might occur.

5. Conclusion

In this article, we introduced recent case studies of problems in digital cordless phones handled by the Technical Assistance and Support Center. With the proliferation of various types of digital cordless phones, the causes of failures are becoming increasingly diverse. The EMC Engineering Group of the Technical Assistance and Support Center aims to achieve prompt resolution of problems related to digital cordless phones and to contribute to the smooth provision of communication services. To this end, it is actively engaged in technology dissemination activities through technical support, development, and technical seminars.

Reference

[1] ARIB RCR STD-28: "Personal Handy Phone System."

External Awards

Research Paper Award

Winner: Kimihiro Mizutani, Takeru Inoue, Toru Mano, and Osamu Akashi, NTT Network Innovation Laboratories; Satoshi Matsuura, Tokyo Institute of Technology; Kazutoshi Fujikawa, Nara Institute of Science and Technology
Date: September 19, 2017
Organization: Japan Society for Software Science and Technology (JSSST)

For "Stable Load Balancing with Overlapping ID-space Management in Range-based Structured Overlay Networks."
Published as: K. Mizutani, T. Inoue, T. Mano, O. Akashi, S. Matsuura, and K. Fujikawa, "Stable Load Balancing with Overlapping ID-space Management in Range-based Structured Overlay Networks," Computer Software, JSSST, Vol. 32, No. 3, pp. 101–110, Aug. 2015.

Best Paper Award

Winner: Takashi Miyamura, NTT Network Service Systems Laboratories; Akira Misawa, Chitose Institute of Science and Technology (CIST); Jun-ichi Kani, NTT Access Network Service Systems Laboratories
Date: September 27, 2017
Organization: The Asia-Pacific Network Operations and Management Symposium (APNOMS) 2017

For "Design of Optical Aggregation Network with Carrier Edge Functions Virtualization."
Published as: T. Miyamura, A. Misawa, and J. Kani, "Design of Optical Aggregation Network with Carrier Edge Functions Virtualization," IEICE Tech. Rep., Vol. 117, No. 186, PN2017-26, pp. 69–74, Aug. 2017.

ACC TOKYO CREATIVITY AWARDS
Interactive Category
New Technology Award

Winner: Katsuhiko Kawazoe, Tomoyuki Kanekiyo, and Shinji Fukatsu, NTT Service Innovation Laboratory Group; Ryuji Kubozono, Akihito Akutsu, Kenichi Minami, Akira Ono, Takashi Miyatake, Koichi Furukado, Hidenobu Nagata, Hiromu Miyashita, Hirokazu Kakinuma, Mariko Yamaguchi, Tetsuya Yamaguchi, Yosihide Tonomura, Takako Sato, Hideaki Takada, Kimitaka Tsutsumi, Yumi Kikuchi, Toshiharu Morizumi, Shingo Kinoshita, Manabu Motegi, Kyoko Hashiguchi, and Kenya Suzuki, NTT Service Evolution Laboratories; Tatsuya Fujii, Naohiro Kimura, Takahiro Yamaguchi, and Ippei Shake, NTT Network Innovation Laboratories; Soichiro Usui, NTT Research and Development Planning Department
Date: September 27, 2017
Organization: All Japan Radio & Television Commercial Confederation (ACC)

For "Kirari! KABUKI."
An initiative for delivering the glamour of kabuki to the world using state-of-the-art technology. A live kabuki performance staged in Las Vegas, USA, was broadcast in Tokyo, Japan. The footage was reproduced later in a highly immersive manner in Kumamoto, Japan.

Spikes Asia 2017
Innovation: Shortlist (Applied Innovation)
Music: Shortlist (Music Live Experience)

Winner: Shingo Kinoshita, Hiroshi Chigira, and Kenichi Minami, NTT Service Evolution Laboratories; Tomoyuki Kanekiyo, NTT Service Innovation Laboratory Group; Kenya Suzuki, Koji Namba, Masato Ono, Motohiro Makiguchi, Hiromu Miyashita, Hirokazu Kakinuma, Takeru Isaka, Hidenobu Nagata, and Hikaru Takenaka, NTT Service Evolution Laboratories
Date: September 29, 2017
Organization: Spikes Asia

For "CYBER TELEPORTATION TOKYO."
The spatial layout of nine transparent screens, including one double-sided screen, was used to create an engaging and immersive experience made possible through a powerful communication infrastructure. The ultralow latency multimedia synchronized transmission technology over high-speed networks achieved a transfer distance of 6543 miles with zero packet loss using 2K x 5 feeds with 600-ms latency (from Tokyo to Austin, Texas, including 75-ms network latency during the event), and 4K x 13 feeds (tested on maximum transfer size in previous trials), and next generation real-time image extraction technology from any background. We applied our technology to extract in 80 ms human shapes of the artists in Tokyo from a background of choice under various stage lighting conditions and without using conventional green screen technology, then transmitted these isolated images. It was the world debut of a double-sided transparent screen projection to display human form image extractions (double-sided screen provided by Polid Screen).

CHEMINAS Outstanding Research Award

Winner: Tetsuhiko Teshima, Hiroshi Nakashima, Yuko Ueno, Satoshi Sasaki, Calum Henderson, and Shingo Tsukada, NTT Basic Research Laboratories
Date: October 5, 2017
Organization: The 36th Chemistry and Micro-Nano Systems conference (CHEMINAS 36)

For "Cell Assembly in Self-folded Multi-layered Soft Films."
Published as: T. Teshima, H. Nakashima, Y. Ueno, S. Sasaki, C. Henderson, and S. Tsukada, "Cell Assembly in Self-folded Multi-layered Soft Films," CHEMINAS 36, 3P-26, Kiryu, Gunma, Japan, Oct. 2017.

Papers Published in Technical Journals and Conference Proceedings

Improved Cladding-pumped 32-core Multicore Fiber Amplifier

S. Jain, T. Mizuno, Y. Jung, A. Isoda, K. Shibahara, J. R. Hayes, Y. Sasaki, K. Takenaga, Y. Miyamoto, S. Alam, and D. J. Richardson

Proc. of the 43rd European Conference on Optical Communication (ECOC 2017), Th.2.D.2, Gothenburg, Sweden, September 2017.

We present an improved cladding pumped high-core-count 32-core multicore-fiber amplifier. A gain of >17 dB and average NF (noise figure) of 6.5 dB with <2 dB variation is obtained for –4 dBm input signal power over all cores in the wavelength range 1534 nm–1561 nm.

Crosstalk Analysis of 32-core Dense Space Division Multiplexed System for Higher Order Modulation Formats Using an Integrated Cladding-pumped Amplifier

C. Castro, S. Jain, Y. Jung, E. De Man, S. Calabrò, K. Pulverer, M. Bohn, J. R. Hayes, S. U. Alam, D. J. Richardson, Y. Sasaki, T. Mizuno, K. Shibahara, T. Kobayashi, Y. Miyamoto, T. Morioka, and W. Rosenkranz

Proc. of ECOC 2017, M.2.E.5, Gothenburg, Sweden, September 2017.

We analyse the crosstalk performance of a fully integrated inline amplified 32-core link for 100G quadrature phase-shift keying (QPSK), 150G 8 quadrature amplitude modulation (QAM), 200G 16-QAM, and 250G 32QAM in a recirculating loop. Transmission distances over 1000 km are confirmed for 8-QAM and QPSK channels.

Hybrid Cladding-pumped EDFA/Raman for SDM Transmission Systems Using Core-by-core Gain Control Scheme

T. Mizuno, A. Isoda, K. Shibahara, H. Ono, M. Fukutoku, and Y. Miyamoto

Proc. of ECOC 2017, M.1.E.3, Gothenburg, Sweden, September 2017.

We propose and demonstrate hybrid cladding-pumped/Raman amplification for multicore fiber transmission systems. 46WDM (wavelength division multiplexing) PDM (polarization-division multiplexed)-16QAM signals were transmitted over a 7-core transmission line and Q-factor enhancement is demonstrated using the core-by-core gain control scheme.

---

50 ch x 250 Gbit/s 32-QAM Transmission over a Fully Integrated 7-core Multicore Link

C. Castro, S. Jain, Y. Jung, E. De Man, S. Calabrò, K. Pulverer, M. Bohn, J. R. Hayes, S. U. Alam, D. J. Richardson, Y. Sasaki, T. Mizuno, Y. Miyamoto, T. Morioka, and W. Rosenkranz

Proc. of ECOC 2017, M.1.E.5, Gothenburg, Sweden, September 2017.

A transmitted distance of 180 km over an integrated multicore link is demonstrated for a C-band 32-QAM WDM system, where the complete usable amplification region of the integrated 7-core amplifiers, supporting 50 channels per core, is exploited.

Adaptive Rates of High-spectral-efficiency WDM/SDM Channels Using PDM-1024-QAM Probabilistic Shaping

H. Hu, M. P. Yankov, F. Da Ros, Y. Amma, Y. Sasaki, T. Mizuno, Y. Miyamoto, M. Galili, S. Forchhammer, L. K. Oxenløwe, and T. Morioka

Proc. of ECOC 2017, Tu.1.D.1, Gothenburg, Sweden, September 2017.

We demonstrate adaptive rates and spectral efficiencies in WDM/SDM (space division multiplexing) transmission using probabilistically shaped PDM-1024-QAM signals, achieving up to 7-Tbit/s data rates per spatial-super-channel and up to 297.8-bit/s/Hz aggregate spectral efficiency using a 30-core fiber on 12.5 and 25-GHz WDM grids.

Recursive Extraction of Modular Structure from Layered Neural Networks Using Variational Bayes Method

C. Watanabe, K. Hiramatsu, and K. Kashino

Proc. of Discovery Science 2017, pp. 207–222, Kyoto, Japan, September 2017.

Deep neural networks have made a substantial contribution to the recognition and prediction of complex data in various fields, such as image processing, speech recognition and bioinformatics. However, it is very difficult to discover knowledge from the inference provided by a neural network, since its internal representation consists of many nonlinear and hierarchical parameters. To solve this problem, an approach has been proposed that extracts a global and simplified structure for a neural network. Although it can successfully detect such a hidden modular structure, its convergence is not sufficiently stable and is vulnerable to the initial parameters. In this paper, we propose a new deep learning algorithm that consists of recursive back propagation, community detection using a variational Bayes, and pruning unnecessary connections. We show that the proposed method can appropriately detect a hidden inference structure and compress a neural network without increasing the generalization error.