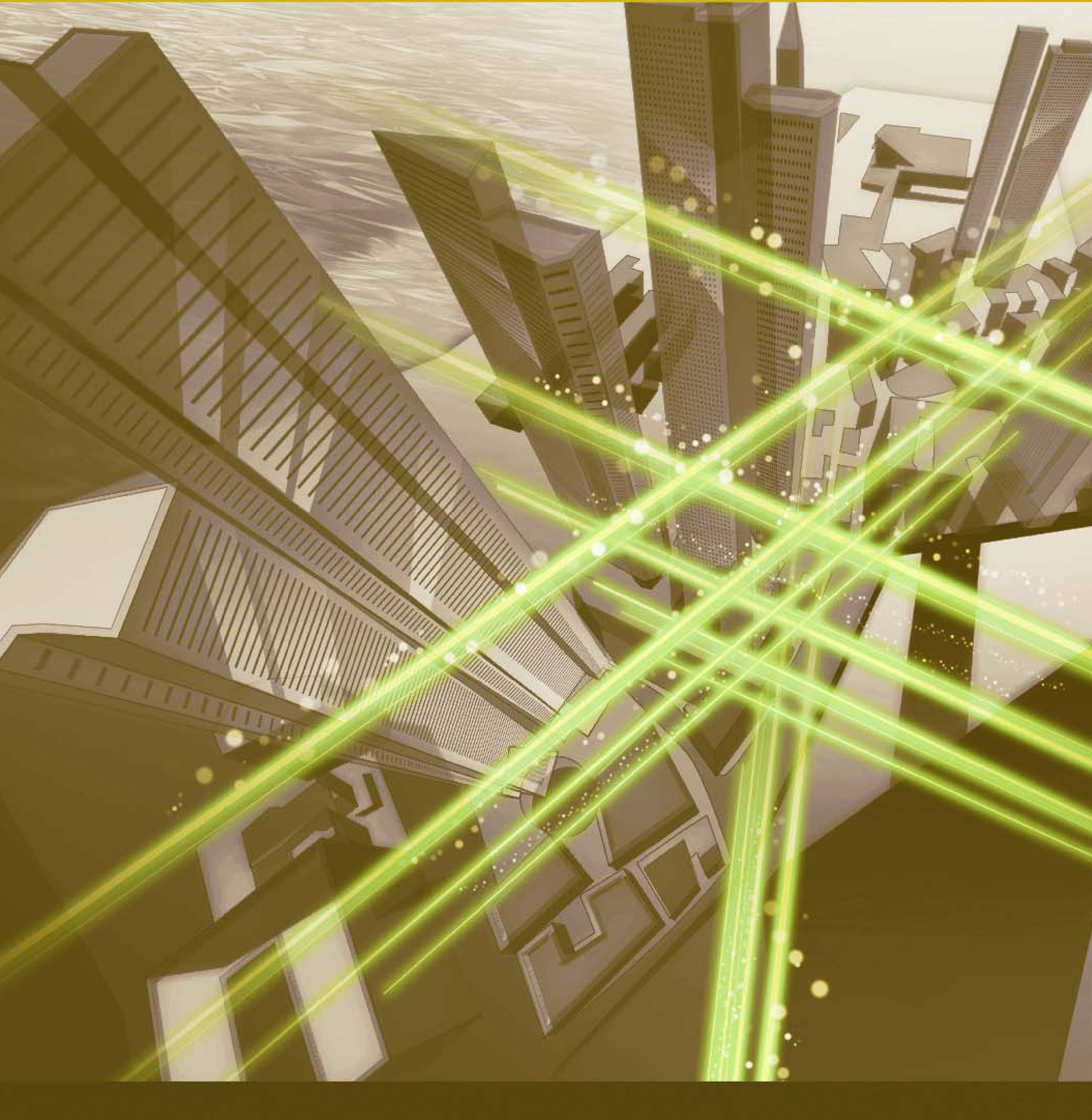
NTT Technical Review

7
2022



July 2022 Vol. 20 No. 7

NTT Technical Review

July 2022 Vol. 20 No. 7

Front-line Researchers

- Ryuichiro Higashinaka, Visiting Senior Distinguished Researcher, NTT Human Informatics Laboratories/NTT Communication Science Laboratories

Rising Researchers

- Takuhiro Kaneko, Distinguished Researcher, NTT Communication Science Laboratories

Feature Articles: Optical and Wireless Transmission Technologies for IOWN/6G

- Research and Development for Pioneering a New Communications Paradigm with Wide-area Coverage
- Future Development of Digital Coherent Optical Transmission Technology
- Research and Development of Scalable Optical Transport Technologies
- R&D Activities of Core Wireless Technologies toward 6G Radio Access

Feature Articles: ICT Platform for Connected Vehicles Created by NTT and Toyota

- Overview of Technical Development and Verification in the Connected-vehicle Field
- Activities and Results of Field Trials—Reference Architecture for a Connected-vehicle Platform
- Activities and Results of Field Trials—Network Edge Computing Platform
- Real-time Spatiotemporal Data-management Technology (Axispot™)
- Selective Vehicle-data-collection Algorithm
- Vertically Distributed Computing Technology
- Lane-specific Traffic-jam-detection Technology
- Technology for Calculating Suddenness Index for Aggregated Values

Regular Articles

- Unsupervised Depth and Bokeh Learning from Natural Images Using Aperture Rendering Generative Adversarial Networks

Global Standardization Activities

- Report of the 9th ITU-T TSAG (Telecommunication Standardization Advisory Group) Meeting

External Awards

Put Your Curiosity First and Try Anything—Researchers Should Do What They Want to Do; Whatever It Takes

Ryuichiro Higashinaka
Visiting Senior Distinguished Researcher, NTT Human Informatics Laboratories/NTT Communication Science Laboratories

Abstract

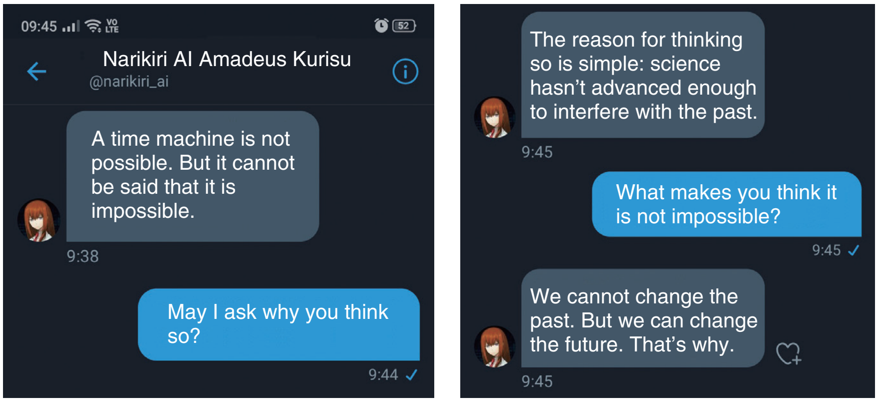
Dialogue systems have made great progress over the last few years due to the application of deep learning and the spread of technologies that enable people to interact with voice agents on smartphones, personal robots, and other devices. Ryuichiro Higashinaka, a visiting senior distinguished researcher of NTT Human Informatics Laboratories/NTT Communication Science Laboratories and professor of the Graduate School of Informatics, Nagoya University, is aiming to create a dialogue system that allows humans and computers to understand each other and intelligently collaborate by clarifying the principles of natural-dialogue interaction between them. We interviewed him about the progress of his research and his attitude as a researcher.

Keywords: dialogue system, common ground, AI

Create a world in which humans and computers can understand each other by pursuing common ground between them

—It has been about three and a half years since the last interview. Can you tell us about the research you are currently working on?

In my last interview, in 2018, I focused on the question-answering system used in the “Shabette Concier” voice-agent service and dialogue systems. Since then, my research has further advanced and expanded into other fields. Natural language processing has rapidly improved through deep learning, and its possible applications have multiplied considerably. For example, the advent of BERT (Bidirectional Encoder Representations from Transformers), a natural-language-processing model, has caused a paradigm shift of sorts and dramatically improved the performance of natural language processing. Over the last few years, I have been applying natural-language-processing techniques based on deep learning to



09:45

Narikiri AI Amadeus Kurisu
@narikiri_ai

A time machine is not possible. But it cannot be said that it is impossible.
9:38

May I ask why you think so?
9:44

The reason for thinking so is simple: science hasn't advanced enough to interfere with the past.
9:45

What makes you think it is not impossible?
9:45

We cannot change the past. But we can change the future. That's why.
9:45

Fig. 1. Narikiri AI Amadeus Kurisu chatbot.

artificial intelligence (AI) systems.

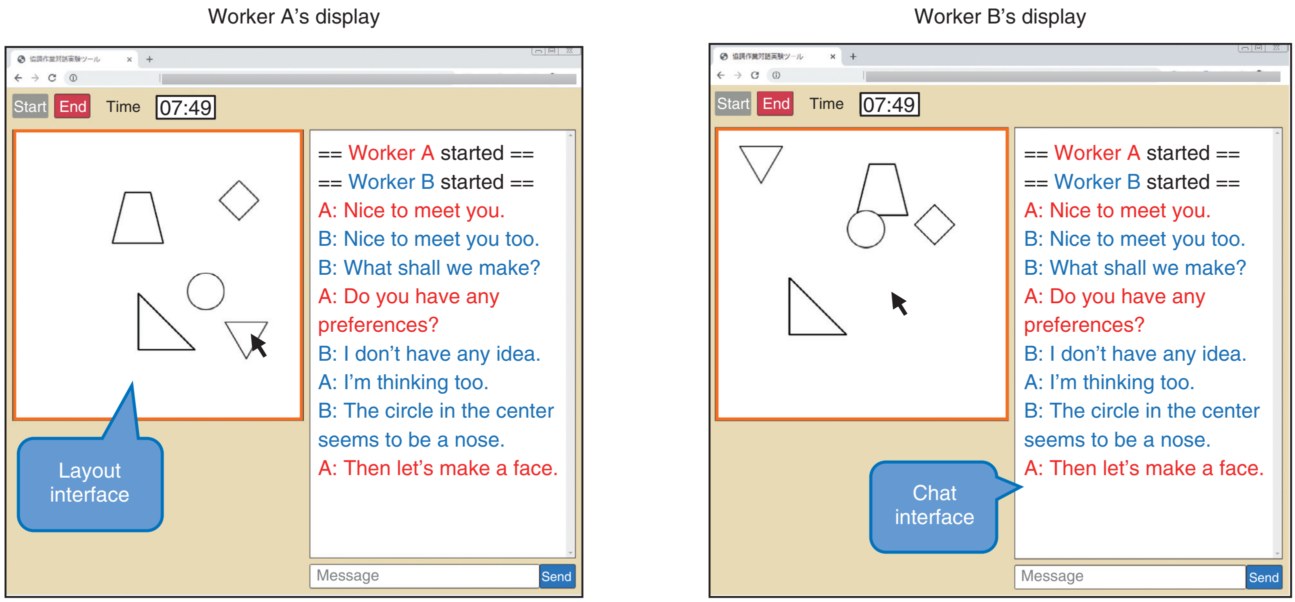
In 2019, as part of the "Todai Robot Project—Can a robot get into the University of Tokyo?" launched by the National Institute of Informatics, we developed an AI robot—using a technology similar to BERT and the latest technology at the time—that takes the written-English subject of the National Center Test for University Admissions and scored a high mark of 185 points (deviation score of 64.1). NTT released a press release about this achievement [1].

We also recently created the chatbot called "Narikiri AI Amadeus Kurisu" (**Fig. 1**), which has greatly improved the accuracy of responses in dialogues compared with the previously introduced "Ayase AI" chatbot. Narikiri AI (*narikiri* means impersonation) uses a unique method of creating a character of a chatbot based on training data obtained from users and fans who describe how a certain character speaks, thinks, etc. We received 45,000 samples for the training data of Narikiri AI Amadeus Kurisu, and by using deep learning, we were able to create a character with high accuracy in response generation. Then, in July 2021, we created an opportunity for fans to interact with the chatbot by using the direct-message feature of Twitter. Although limited to three days, the event was very well received; in fact, some participants had hundreds of conversations with the chatbot, and the event received a high rating of 4.59 out of 5 in a survey. My paper on Narikiri AI was accepted by a top conference on language processing [2], so I'm proud that this work has been recognized as a significant academic achievement. However, some of the chatbot's responses were not suitable for the character, so we are currently analyzing user feedback and improving the chatbot.

—The conversation between computers and humans seems to be evolving. Have we reached a level at which computers and humans can mutually understand each other?

I'm afraid we haven't reached that level yet. To enable computers and humans to understand each other and communicate, we need to develop a system that can build *common ground* between them. Common ground is the content that is understood by the participants in conversation with each other. I believe that once common ground between computers and humans can be established, mutual understanding can be increased and communication between them will evolve rapidly, making collaboration between dialogue systems and humans possible. Although common ground is a concept that has been discussed in textbooks for more than 20 years, it has not been engineered or implemented in a system. However, I decided to pursue this theme because I believe that if common ground cannot be established, dialogue systems won't progress, and the next generation of such systems won't be viable.

Common ground is invisible because it is in the mind, which made research on common ground difficult. To visualize the process of establishing common ground, we set up a task called *CommonLayout* in which each of the two participants views their own figure-arrangement screen and cooperatively determines the placement of figures while interacting with each other via text chat and collected large-scale dialogue data between people executing this task (**Fig. 2**) [3]. We quantified common ground by constantly recording the placements of the figures and measuring the distance between each figure placement



== Worker A started ==
== Worker B started ==
A: Nice to meet you.
B: Nice to meet you too.
B: What shall we make?
A: Do you have any preferences?
B: I don't have any idea.
A: I'm thinking too.
B: The circle in the center seems to be a nose.
A: Then let's make a face.

== Worker A started ==
== Worker B started ==
A: Nice to meet you.
B: Nice to meet you too.
B: What shall we make?
A: Do you have any preferences?
B: I don't have any idea.
A: I'm thinking too.
B: The circle in the center seems to be a nose.
A: Then let's make a face.

Fig. 2. CommonLayout task.

(the sum of vector differences defined between any two figures) during the task. In other words, we made it possible to visualize the degree to which common ground was being established at each point during the dialogue. By analyzing these data, we found that the process of establishing common ground can be categorized into several clusters. This research was presented at the annual conference of the Association for Natural Language Processing and received an award for excellence in March 2021. In the process of analyzing these data, we also learned that so-called *naming* plays a key role in building common ground. In particular, humans seem to build common ground by naming figures and matching their perceptions of those figures with those of others [4].

In addition to the above-mentioned work, to clarify the effect of modality (channels of information such as text, audio, and video) and social relationships on the process of constructing common ground, I'm extending my research on the CommonLayout task conducted via text chat to construct and analyze a corpus (a large-scale collection of natural language and other data) composed of audio and video. We found that it is easier to establish common ground when audio and video are used rather than when only audio is used. We also found that it was easier to establish common ground when the participants were acquainted rather than when they met for the first time. The reader might think these findings are obvious. However, the process of building common ground differs according to the conditions, and I believe that these differences will lead to important findings. By analyzing the effects of modality and social relationships on the establishment of common ground, we hope to create a dialogue robot that can smoothly establish common ground between computers and humans. I'd also like to establish common ground from various angles, and we are working on the task of naming the shapes of the pieces in a tangram (a puzzle composed of seven pieces dividing a square) while the participants are having a dialogue during the task.

I want to accelerate my research and expand it by combining different disciplines

—How much research on common ground has been done?

Although researchers in the philosophy of language have examined how humans establish common ground, engineering approaches are few and far between; in fact, only a few researchers around the world are focusing on this theme. I'm keeping my eyes open and reviewing the literature, but there are currently not many relevant reported studies. Therefore, to create a future in which AI and humans can collaborate, I'm accelerating research on this theme with determination. I'm also promoting joint research. For example, the above-mentioned tangram-naming task is pursued in a joint research with Shizuoka University, and I'm collaborating with Professor Yugo Takeuchi, an expert in cognitive science.

I'm also working with Professor Yasuhiro Minami of the University of Electro-Communications on researching the naming in building common ground I mentioned earlier. In addition to studying speech processing, Professor Minami is also studying language development in young children. Through our collaboration, we are thus able to pursue engineering and language development in combination. I'm also collaborating with Professor Kazunori Takashio, an expert in social robotics at Keio University, on the effect of the modality and social relationships in building common ground. Through these collaborations, I'm advancing my research on common ground by combining various disciplines.

—You seem to be engaged in academically and socially significant activities, including active joint research.

I think there is academic significance in doing what has been avoided so far. When researching dialogue systems, it is relatively easy to analyze the verbal exchanges that are visible as text and manifested as phenomena. However, it is difficult to analyze and evaluate what the other party is thinking when they say those words. The research on the CommonLayout task is to visualize those thoughts and eliminate that difficulty. My research is progressing, and I'm confident that the day is near when we will be able to build a dialogue system that enables humans to work together with AI when thinking about and deciding on product placement and other issues.

I think that research on dialogue systems is currently at a standstill. That is to say, deep learning is still limited in what it can do. We can have a conversation with an AI system, but it is still only for a short time. If the system forgets yesterday's communication, intellectual collaboration between humans and AI will not be established. I want to develop AI capable of long-term communication to be a partner for us humans so that it can accompany us through our lifetime.

I believe that by pursuing common ground, we can make human society better 20 years or so from now. As I say in my book "Chatting Skills of AI" (KADOKAWA), I hope to create a world in which humans and computers can build a two-way relationship and enhance mutual understanding.

Be willing to go beyond your area of expertise

—What qualities are required of researchers today?

Compared with 2001, when I joined NTT, it is now possible to conduct many experiments relatively inexpensively through crowdsourcing and other means. Therefore, I think that researchers are in a fortunate environment. On the contrary, companies with vast financial and human resources, such as GAFA (Google, Apple, Facebook, and Amazon), are investing in research and development to rapidly bring new products and services to the world. The improved research environment has increased the tempo of research and development and intensified competition.

To survive in these times, researchers are required to constantly take on new challenges. It is also necessary to keep in mind our impact on—and contribution to—society as we pursue our research. To meet these requirements, I believe it is important to be willing to go beyond our own area of expertise. Many problems can only be solved by incorporating knowledge from other research areas, and we are no longer in an era of sticking to only one area of specialization such as language processing; rather, we are required to have a multifaceted and broad perspective. For example, when building a dialogue system, we must keep in mind the development of laws, etc. and consider society, users, etc. in a composite manner; otherwise, even if we develop a system born of outstanding research, we will not be able to roll it out.

In fact, dialogue systems were not the theme that interested me when I joined the company. I wanted to pursue machine translation because I loved words, but the theme assigned to me was dialogue, which I found so interesting that I ended up doing it for 20 years. Since that experience, I've been working with the spirit of "let's try everything." Basically, I never refuse a request.

—What would you like to say to young researchers?

I believe that researchers should do what they want to do; whatever it takes. I also believe one of the important roles of a senior distinguished researcher is to set an example and show that one needs to do what one wants to do. It is important to be interested in new things, have a multifaceted and broad perspective beyond one's area of expertise, and not be afraid to try anything. Why not start by putting your curiosity first and trying anything? When we are young, we sometimes think that one mistake means it's all over. Immediately after I joined NTT, I was worried that if I did not get a paper accepted by a top conference within about three years, I'd be on my way out. Fortunately, I managed to get a paper accepted with the help of my mentor; even so, I was still afraid of failure and thought I might be branded an unfit researcher.

However, I don't want young researchers to worry about failing. That's because they can use their failures to move on to the next stage of their development. To put it another way, if you think of years of service as the denominator and failure as the numerator, the denominator is small when you are young, so the damage of failure may be severe, but as the denominator becomes larger, that severity is diluted by your longer service, so the way you perceive failure changes.

As a member of a corporate organization, you will sometimes be obliged to follow company policy and instructions from supervisors. However, it is important not to abandon what you want to do to focus on your obligation but to connect with peers as well as with researchers who share similar interests. In my case, the connections that I've made with the academic community have been a great help. It is a good idea to build relationships not only with people inside your company but also with people outside the company from a young age.

References

[1] Press release issued by NTT, "AI Achieved a Score of 185 on the English Written Exam of the National Center Test for University Admissions in 2019," Nov. 18, 2019.
https://group.ntt/en/newsrelease/2019/11/18/191118a.html

[2] K. Mitsuda, R. Higashinaka, T. Li, and S. Yoshida, "Investigating Person-specific Errors in Chat-oriented Dialogue Systems," Proc. of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 464–469, Dublin, Ireland, May 2022.

[3] K. Mitsuda, R. Higashinaka, Y. Oga, and S. Yoshida, "Dialogue Collection for Recording the Process of Building Common Ground in a Collaborative Task," Proc. of the 13th Language Resources and Evaluation Conference, pp. 5749–5758, Marseille, France, June 2022.

[4] Y. Saito, K. Mitsuda, R. Higashinaka, and Y. Minami, "Analysis of Naming in the Process of Building Common Ground in Dialogue," Proc. of the 28th Annual Meeting of the Association for Natural Language Processing, pp. 38–42, Mar. 2022 (in Japanese).

■ Interviewee profile

Ryuichiro Higashinaka received a B.A. in environmental information, Master of Media and Governance, and Ph.D. from Keio University, Kanagawa, in 1999, 2001, and 2008. He joined NTT in 2001. Since 2020, he has been a professor at the Graduate School of Informatics, Nagoya University and visiting senior distinguished researcher at NTT. His research interests include building question-answering systems and spoken dialogue systems. From November 2004 to March 2006, he was a visiting researcher at the University of Sheffield in the UK. He received the Maejima Hisoka Award from the Tsushinbunka Association in 2014 and the Prize for Science and Technology of the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology in 2016. He is a member of the Institute of Electronics, Information and Communication Engineers, the Japanese Society for Artificial Intelligence, the Information Processing Society of Japan, and the Association for Natural Language Processing.

Learning 3D Information from 2D Images Using Aperture Rendering Generative Adversarial Networks toward Developing a Computer that "Understands the 3D World"

Takuhiro Kaneko
Distinguished Researcher, NTT
Communication Science Laboratories

Abstract

When people look at photos, they can estimate three-dimensional (3D) information, such as depth, from their experience and knowledge, but computers have difficulty in doing so because they cannot have such experience and knowledge. We spoke to Takuhiro Kaneko, a distinguished researcher who developed a novel deep learning model that can learn 3D information from standard 2D images such as those on the web.

Keywords: generative adversarial networks, unsupervised learning, depth and bokeh

Learning of depth and bokeh effects from natural images with aperture rendering generative adversarial networks

—What is research on "learning of depth and bokeh effects from natural images" about?

It is difficult to record the three-dimensional (3D) world in which we live in as it is. For this reason, it is common to record and store two-dimensional (2D) images such as photographs instead of 3D information.

When people look at photos, they can estimate 3D information, such as depth, from their previous experience and knowledge. However, computers have difficulty in doing so because they do not have such experience or knowledge. In the future, when we think about scenarios where robots support our lives, it will become essential for computers to understand the 3D world. The easiest way for computers to learn is to provide a large number of pairs of 2D images and 3D information as training data. Learning would be easy because they know the correct answer. However, this method requires special devices such as depth sensors and stereo cameras and is costly.

For this reason, in this research, we created a deep learning model that can learn 3D information from standard 2D images such as those taken with ordinary

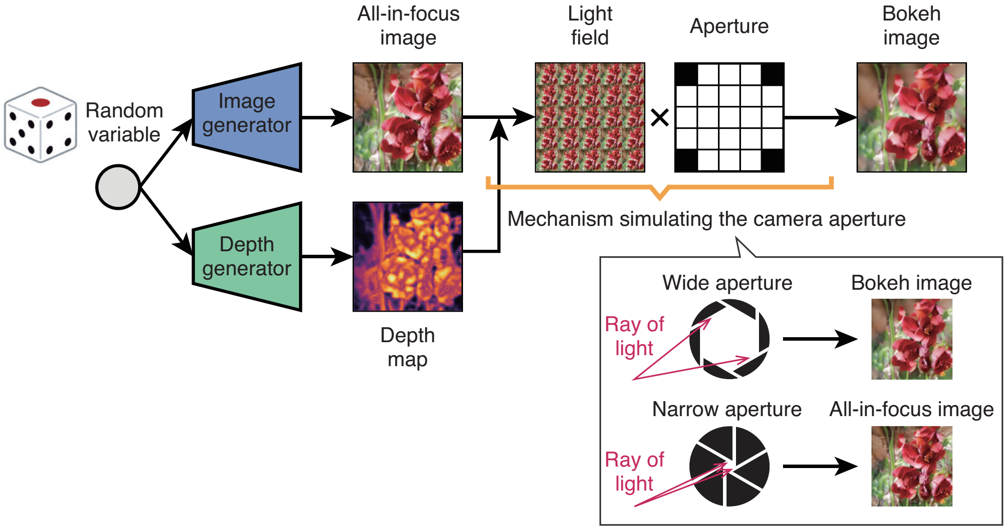


Fig. 1. Process flow in the AR-GAN generator.

cameras and those on the web. When we look at the photos, we see that the focus is usually on the object, and the background is often blurry. Using these defocus cues as clues, we carry out learning of 3D information, especially depth and bokeh effects. In other words, if the problem of projecting 3D information onto 2D images through a camera is the forward problem, for this research, the goal is to solve the inverse problem. This problem is challenging because it is the so-called “ill-posed problem” of estimating 3D information from only 2D images that lack various information.

—Specifically, through what mechanisms are you carrying out learning?

This research is based on a technology called a generative adversarial network (GAN). GAN is a so-called “unsupervised learning model,” which does not require pre-determined correct answers, and consists of two neural networks; namely, a generator and discriminator. The generator generates a “fake image” from a given random variable. On the other hand, the discriminator distinguishes two types of images, “real images” and the “fake images” generated by the generator. Because they work adversarially, wherein the generator tries to trick the discriminator while the discriminator tries to identify the fake images precisely, learning can be carried out while competing, and as a result, the generator can produce realistic images.

Since GAN is a specialized technology for producing 2D images, it has no connection with the 3D world. We, therefore, proposed the “aperture rendering GAN (AR-GAN),” which incorporates the optical properties of the camera into the GAN. “Aperture” refers to the aperture of the camera. By incorporating optical constraints due to the aperture of the camera when projecting the 3D world into 2D images, the generator can learn by associating 2D images, depth maps, and bokeh effects.

Figure 1 shows the process flow in the AR-GAN generator. The “image generator” is also included in the GAN and generates an all-in-focus image when given a random variable. The depth generator generates the depth map that is paired with the image, and is unique to AR-GAN. The system then uses these pairs of generated data to enable a mechanism simulating the camera aperture. The light field consists of 25 images, which represent what the image looks like in the aperture of the camera. In the center, the image created by light coming in through the center of the aperture is displayed. When the object is offset from the center to the right, an image of the object viewed from the right part of the aperture is displayed; and when the object is offset to the top of the aperture, an image of the object viewed from the top part of the aperture is displayed.

It may be easier to imagine the mechanism by moving your face up, down, left, and right while looking at the object in front of your eyes. Objects with focus are clearly visible because they are in the same position and do not move when you move your face, but objects that are farther away move more when you move your face. Therefore, summation of these images produces images with bokeh effects. This is how the mechanism enables synthesizing a bokeh image on the basis of the predicted depth map and all-in-focus image using a camera with an optical constraint on the light field.

—How is the progress of the research and what are the challenges going forward?

Currently, we are able to generate flower images, bird images, and human face images. Learning can take days, but once it is done, the system can generate in a few seconds bokeh images that cannot be distinguished from the real images. At present, I think that if we narrow down the types of objects, we would be able to generate reasonable images.

As for the challenges, one is that depending on the type of object, some are easier to learn while others are not as amenable to learning. For example, for human face images, the shape and position of the parts are fixed to some degree, making it easy to learn them. On the other hand, images with many variations, such as those of animals taken from different viewpoints and distances, are difficult to learn.

Also, the larger the image size, the more details need to be synthesized, and the more processing time it takes, making learning more difficult. I think there will be various issues that will emerge as we expand the scope of applications in the future. As expected, the difficulties inherent with ill-posed problems remain.

Developing a computer that understands the 3D world

—What will this technology enable in the future?

Our mission as researchers is to “develop computers that are highly compatible with humans.” For this reason, it is essential to understand the 3D world; but the cost of data collection is likely to be a barrier to applying the technology to a wider range of fields. In that regard, I think this research is beneficial because it is excellent in terms of data collection. In the future, we will be able to create robots that can move freely around the 3D world, build 3D worlds without dissonance in virtual space, and develop tools to create 3D objects in virtual space.

Another advantage is that optimized models can be built as long as we can collect 2D images. For example, if you gather photos taken by a well-known photographer, you can build a model that learns the bokeh effects unique to that photographer. Currently, communication using images such as through social media has become very common. If we can easily add the bokeh effects, it may become easier to create more attractive photos. I think this is particularly useful in the three areas of robotics, content generation, and entertainment.

—What are the future prospects and initiatives on collaborations with other fields?

Since it is basic research, it is difficult to set specific targets for practical use at this time; but we plan to continue to improve performance by increasing accuracy and resolution. We have focused on the camera aperture, but it is interesting to note that adding physical constraints enables creating a more reliable computer.

One of the key technical areas of the Innovative Optical and Wireless Network (IOWN) initiative is “Digital Twin Computing,” where computations are performed using digital twins of various industries, things, and humans. In order to merge the real world with the virtual world, it is necessary for computers to properly understand the real world, so I believe that this technology can also contribute to these areas.

I feel that the importance of integrating media generation technology with various fields is increasing as the technology matures. Currently, we are conducting research that combines computer science, such as computer vision and machine learning, with physics, such as optics. Going forward, we will continue to focus our efforts on collaborations with people from other fields, such as computer graphics for image creation and photonics for photography, as well as on cross-discipline implementation.

—Could you give a message to young researchers and future business partners?

NTT laboratories are conducting extensive research ranging from basic research to applications, and in particular, our laboratory, NTT Communication Science Laboratories, is conducting research on how to improve communication between humans and between humans and machines. In the past, we have often been limited to basic research, but recently, the distance between basic research and applied research has been narrowing, and I feel that the opportunities for engaging in research while considering real-world problems are increasing. We are also seeing more and more cases of the technology presented at international conferences being embedded in applications and deployed as a service on the web. I think it is interesting to see the output, for example, by creating a solution that actually converts voice, rather than stopping at tinkering with formulas.

There are limitations to doing research individually. Our laboratory is also actively collaborating with universities and accepting interns, so I hope that we can continue to actively collaborate particularly with students and young researchers who want to create something or change something.

As for business partners, since we have been able to create interesting ideas and technologies from the research side, we need their help in linking these ideas and technologies to services. On the other hand, we can get ideas for research by receiving feedback from people in the service field who have profound knowledge of real problems. So, going forward, we hope to continue to actively collaborate also with business partners.

■ Interviewee profile

Takuhiro Kaneko received his M.E. degree from the University of Tokyo in 2014. He joined NTT the same year as a member of NTT Communication Science Laboratories. In 2020, he completed his Ph.D. at the University of Tokyo. He has been a distinguished researcher at NTT Communication Science Laboratories since 2020. He is engaged in research on computer vision, signal processing, machine learning, and deep learning, particularly dealing with image synthesis, speech synthesis, and voice conversion. He received the Hatakeyama Award from the Japan Society of Mechanical Engineers, the ICPR Best Student Paper Award in the 21st International Conference on Pattern Recognition (ICPR 2012), and the Dean’s Award for Best Doctoral Thesis from the University of Tokyo. For details: https://www.kecl.ntt.co.jp/people/kaneko.takuhiro/

Research and Development for Pioneering a New Communications Paradigm with Wide-area Coverage

Shigeru Iwashina, Katsuhiro Shimano, Koichi Takasugi, Kazunori Akabane, and Takashi Saida

Abstract

Research and development at NTT Network Innovation Laboratories aims to establish elemental technologies for next-generation communication networks envisioned under NTT's Innovative Optical and Wireless Network (IOWN) and the 6th-generation mobile communication system (6G). These technologies, which include advanced and high-capacity backbone optical transmission networks and extended coverage of wireless communications, are being developed to support dramatic changes in the information society such as the expansion of remote work due to the COVID-19 pandemic. This article introduces optical/wireless transmission technologies and systemization technologies currently being researched and developed at NTT Network Innovation Laboratories.

Keywords: optical transmission technology, wireless transmission technology, IOWN/6G

1. Introduction

Social activities that enable remote participation have become widespread as reflected by the adoption of remote work as a countermeasure against the COVID-19 pandemic. To support these activities, there are growing expectations for the development of next-generation communication networks envisioned under NTT's Innovative Optical and Wireless Network (IOWN) and the 6th-generation mobile communication system (6G) featuring, for example, advanced and high-capacity backbone optical transmission networks and extended coverage of wireless communications. Research and development (R&D) at NTT Network Innovation Laboratories aims to establish communication technologies that can exploit the physical waves of diverse frequency bands including light, radio, and sound waves and enable long-distance, high-speed, and high-capacity transmission in a wide range of areas including communication media such as optical fiber, water, and air. Researchers at NTT Network Innovation Laboratories understand that their mission is to pioneer a new communications paradigm by achieving the above research targets using scientific knowledge based on physics and mathematics such as that related to electromagnetic wave propagation and light propagation, digital signal processing, and media processing.

As shown in **Fig. 1**, the research scope of NTT Network Innovation Laboratories covers a wide range of areas. Researchers usually dealt with optical fiber and radio wave propagation, but today, the research scope has broadened as far as low-frequency bands, such as sound waves, and research targets have extended beyond air and enclosed spaces, such as optical fiber, to include all types of spaces and media from water to outer space as communication media. Research in underwater acoustic communication targeting water, for example, aims to achieve information transfers beyond the Mbit/s class capable of video transmission. To achieve extreme-high-capacity wireless transmission toward 6G [1], NTT Network Innovation Laboratories is also researching and developing new areas such as orbital angular momentum (OAM)

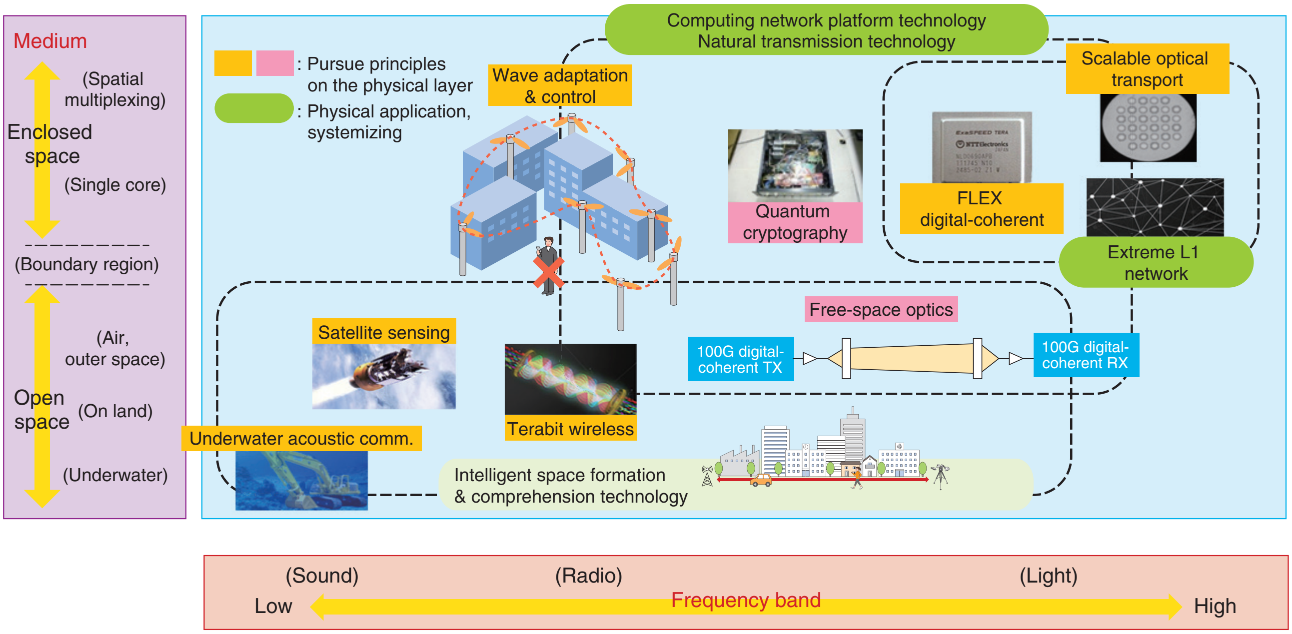


Fig. 1. Research scope of NTT Network Innovation Laboratories toward IOWN/6G.

multiplexing transmission wireless communications, free space optics (FSO) assuming air and outer space as media, and quantum cryptography communications for achieving the ultimate in security.

At NTT Network Innovation Laboratories, we are also engaged in specific R&D toward the development of the All-Photonics Network (APN) [2], which is one of the three main pillars of NTT's IOWN initiative. This includes the R&D of elemental technologies for achieving a next-generation optical communication network that can easily provide high-capacity optical communication paths to a variety of customers as expected of the APN. Specific projects include the development of a low-power digital signal processor (DSP) for achieving 400-Gbit/s-class signal transmission through digital-coherent optical transmission technology, demonstration of bulk extension of the optical amplification bandwidth through wideband optical parametric amplifiers, and the development of technology for automatically setting high-capacity optical communication paths.

The remainder of this article introduces a number of initiatives in cutting-edge technologies undertaken by NTT Network Innovation Laboratories toward IOWN/6G divided into frontier communication technologies, wave propagation technologies, and innovative transport technologies.

2. Frontier communication technologies

An outline of frontier communication technologies is shown in **Fig. 2**. With the aim of achieving cyber-physical systems such as smart cities, we are working on optical-path automatic optimization/control technology for connecting computing resources distributed over a wide area on the IOWN APN and high-speed remote data-transfer technology for achieving high-capacity, low-latency data transfers. These technologies will lead to the establishment of an optical network platform for ultra-wide-area distributed computing. We are also working on natural media processing technology to transmit information on the configuration of physical space in real time with low latency toward ultra-high-presence communication including video and on multimodal wireless environment-comprehension/prediction technologies to predict, set, control, and manage communication quality and the environment through sensing and artificial intelligence (AI). We are also taking up the challenge of establishing secure-transmission technology for envisioning the use of quantum computers.

2.1 Optical network platform for ultra-wide-area distributed computing

At NTT Network Innovation Laboratories, we are developing technology for dynamically connecting

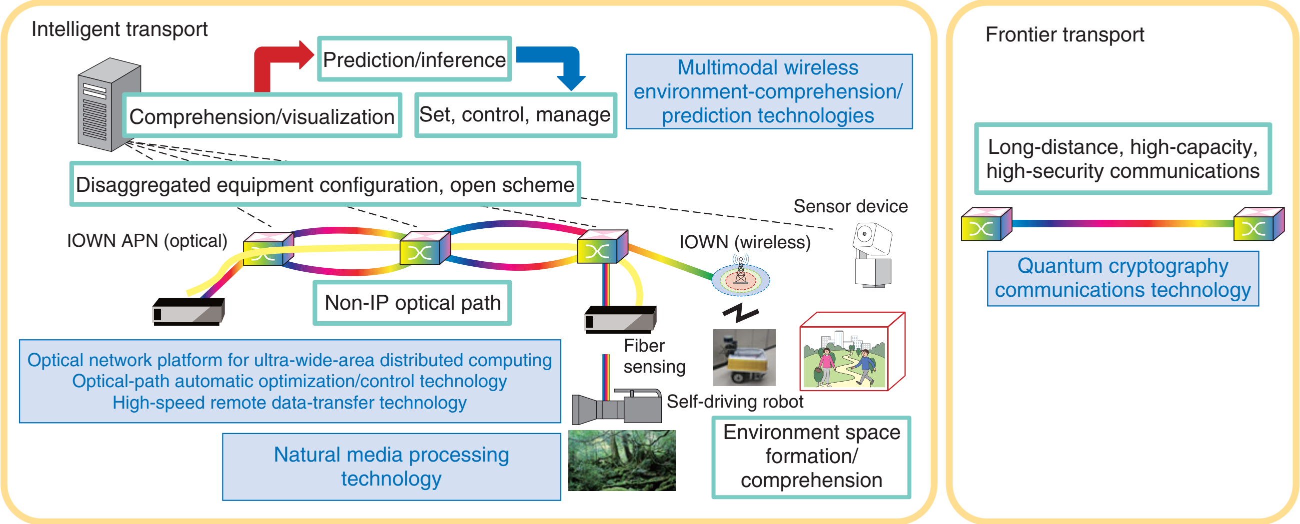


Fig. 2. Outline of frontier communication technologies.

computing resources, such as accelerators and memory devices, distributed over an ultra-wide area on the APN through the optimal design and control of transmission modes and optical paths on the basis of optical characteristics. This technology will enable the construction of optical paths that extract maximum optical transmission performance. Our aim is the establishment of an optical network platform toward ultra-wide-area distributed computing essential to the development of smart cities through high-speed remote data-transfer technology, which will make exclusive use of bandwidth on ultra-long-distance optical paths and synchronize remote memory devices with extreme low latency.

2.2 Natural media processing technology

We are developing spatial reconstruction technology with extreme low latency with a view to ultra-high-presence communication. This will be accomplished by accommodating information on the configuration of physical space in APN optical paths as a time series of high-speed digital signals, enhancing high-speed media transfers with low-latency features to enable long-distance transmissions with extreme low latency, and using lightweight machine learning. Looking to the future, we aim to establish cross-layer AI monitoring technology to collect the input/output of people, things, and the environment across multiple layers and make inferences from those data. Our goal with this technology is to enable automatic optimization even during system operation through state monitoring and detection of anomalies and their causes.

2.3 Multimodal wireless environment-comprehension/prediction technologies

To satisfy a variety of requests for network services to be provided by IOWN, we are developing technologies for using diverse types of physical space information obtained using cameras and sensors for predicting communications quality several seconds into the future and for maintaining high-quality communications at all times through optimal communication means. Our goal is to establish environment-comprehension technology that uses the relationship between physical space information and wireless communication system information to extract the former, such as location information, from the latter.

2.4 Quantum cryptography communications technology

NTT Network Innovation Laboratories is exploring new technology areas based on optical technologies toward long-term innovative contributions to the IOWN initiative. Specifically, we aim to use optical-fiber transmission technology and communication protocol technology to achieve a quantum cryptography communications system that overcomes the limitations in distance and communication speeds in conventional quantum key distribution. We will also endeavor to establish security and transmission technologies for the age of quantum computers.

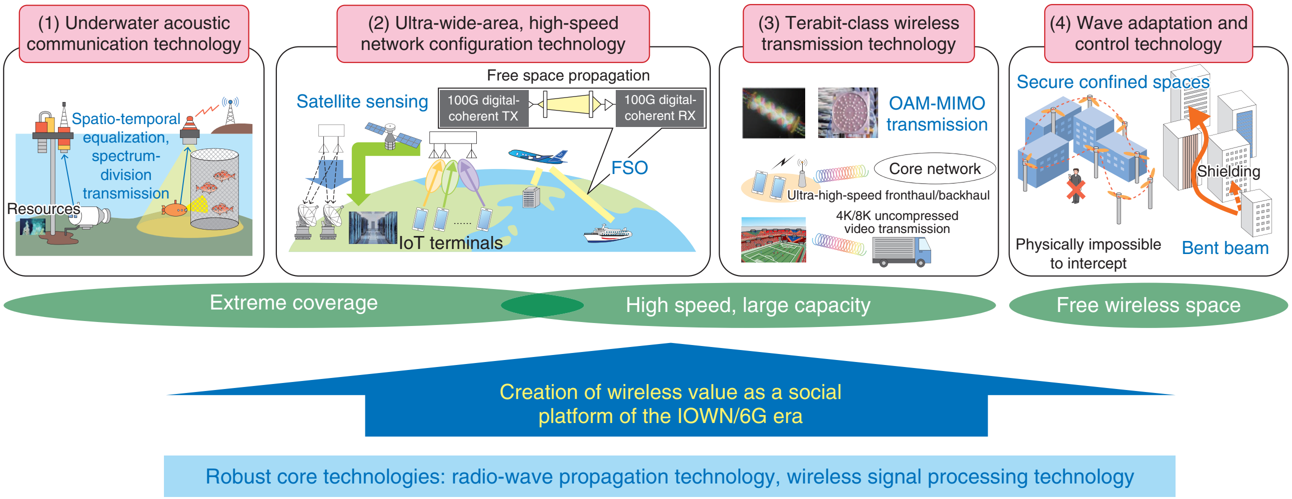


Fig. 3. R&D related to wave propagation technology.

3. Wave propagation technologies

Wireless communications in the IOWN/6G era is expected to have a variety of features. These include extreme coverage extension that includes coverage in the sea and outer space in addition to terrestrial communications, high-speed/high-capacity transmission irrespective of the points being connected, and flexible wireless space formation that confines radio waves to the targeted area, all with the aim of creating new value in wireless communications. At NTT Network Innovation Laboratories, we are working on four key technologies toward the creation of wireless value as a social platform for the IOWN/6G era: (1) underwater acoustic communication technology, (2) ultra-wide-area, high-speed network configuration technology, (3) terabit-class wireless transmission technology, and (4) wave adaptation and control technology (**Fig. 3**).

3.1 Underwater acoustic communication technology

Wireless communication technology based on radio waves capable of high-speed, stable, and long-distance communication has not been established for underwater applications; as a result, unmanned probes and undersea heavy equipment used for resource development on the ocean floor or for port construction typically connect with a support vessel on the surface of the sea using a communication cable 100 m or longer to achieve remote operations [3]. This state of affairs has been a major issue from the viewpoint of work efficiency. We aim to establish Mbit/s-class underwater acoustic communication technology to enable wireless remote control of underwater equipment and devices. This technology will include spatio-temporal equalization technology to compensate for significant waveform distortion of acoustic waves and spectrum-division transmission technology to achieve high speeds through broadband transmission.

3.2 Ultra-wide-area, high-speed network configuration technology

Areas in which people do not live, such as in highly rural and mountainous areas and the open sea, may not have a terrestrial wireless communication infrastructure for economic reasons. Our aim is to achieve ultra-wide-area, high-speed networks in areas where no wireless communication infrastructure has been set up. To this end, we are pursuing satellite Internet of Things (IoT) platform technology [4] to enable the collection of sensor data on a global scale and FSO platform technology to enable 100-Gbit/s-class high-speed communications in areas such as the open sea.

3.3 Terabit-class wireless transmission technology

Terabit-class wireless transmission technology is considered necessary for the fronthaul and backhaul in the IOWN/6G era [5]. We aim to further increase transmission bandwidth and the amount of spatial multiplexing essential to terabit-class wireless transmission through digital signal processing technologies.

These include OAM and line-of-sight multi-input multi-output (MIMO) schemes for achieving spatial multiplexing and high-frequency-band Butler circuit configuration technology for increasing the bandwidth of OAM multiplexed transmission.

3.4 Wave adaptation and control technology

In wireless communications, radio waves tend to spread in all directions even outside the area targeted for communications. This leakage of radio signals outside the intended area results in a drop in confidentiality, increase in interference, and power loss as universal problems in wireless communications. We aim to establish the ultimate in wireless communications by preventing the leakage of radio waves using wave-control techniques. These will include terminal-coordinated, user-centric radio access network technology that links not only base stations but also terminals to form wireless space in an adaptive manner and multi-shape wave-control technology to fully control radio-wave trajectories.

4. Innovative transport technologies

NTT Network Innovation Laboratories is pursuing innovative transport technologies for achieving extreme-high-capacity optical paths as a platform for IOWN. In particular, we are researching and developing digital-coherent optical transmission technology including the development of a low-power, 1-Tbit/s-class DSP for optical communications as well as opto-electronic integration technology to link light and electricity and analog and digital circuits. We are also working on innovative Layer 1 networking technology to generate new value for users and operators and pioneering elemental technologies toward a dramatically new user experience (UX) in the APN. With the aim of efficiently accommodating massive amounts of traffic in the future, we have taken on the development of innovative extreme-high-capacity transmission technology with a 10-Pbit/s-class throughput per node, which is enabled by scalable optical transport platform technology driving the development of optical signal processing technology.

4.1 FLEX digital-coherent optical transmission platform technology

NTT Network Innovation Laboratories aims to establish FLEX digital-coherent optical transmission platform technology as an elemental technology for economically achieving IOWN (**Fig. 4**). This technology will achieve long-distance and low-power high-capacity optical transmission of the 1-Tbit/s class essential to the construction of the APN. In addition to conventional long-distance transmission oriented to carriers, this technology is also applicable to short-range transmission such as datacenter interconnects. For rapidly expanding application areas, it aims to achieve optimal optical paths by using a digital signal processing function to make flexible changes to the transmission scheme, type of compensation processing, etc. in addition to high-accuracy transmission path monitoring.

4.2 Scalable optical transport platform technology

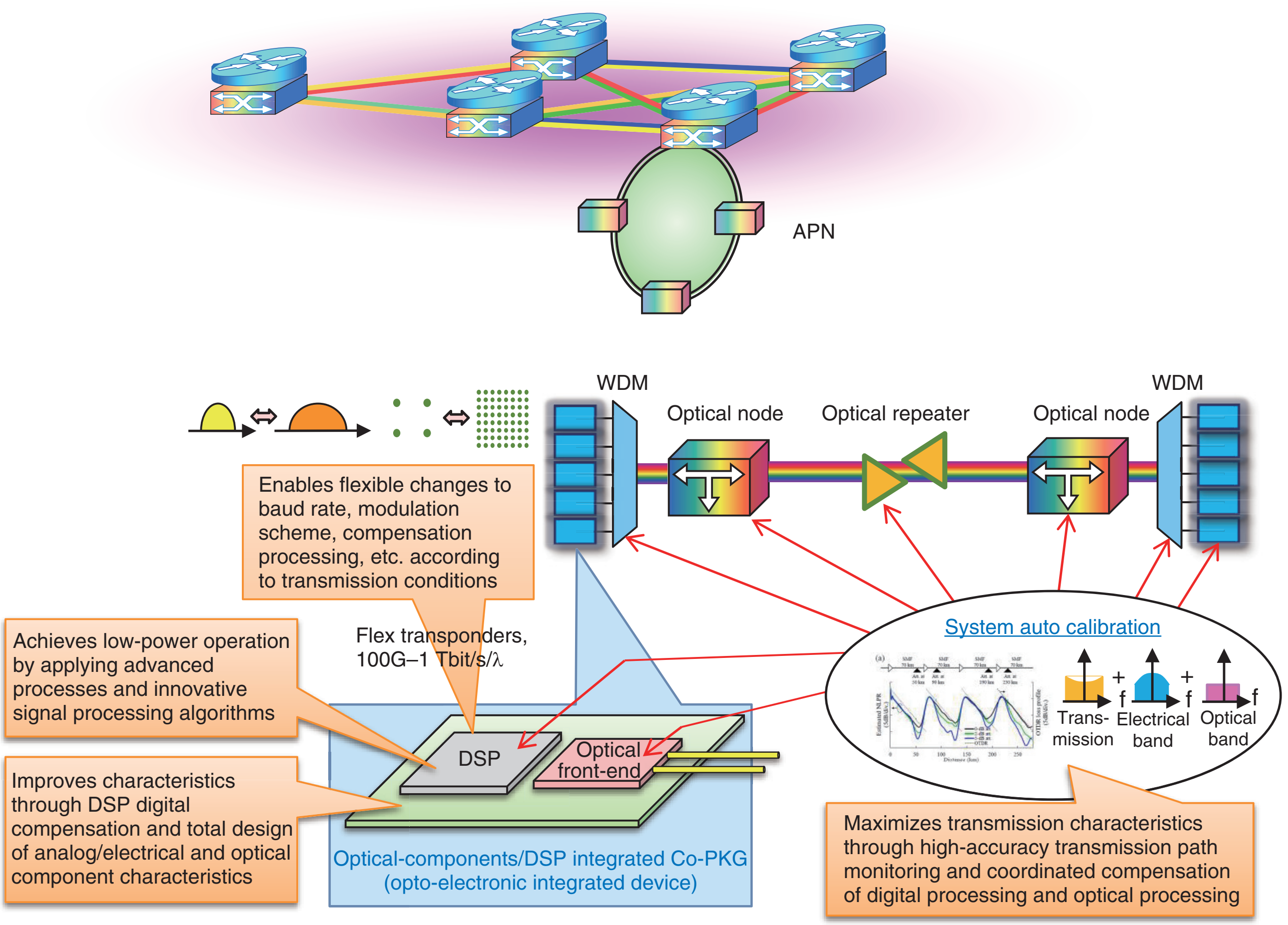
A petabit-class optical network will be needed to accommodate the massive volume of communications traffic that will be generated by the expansion of cloud services and further penetration of smartphones. As part of our efforts to meet this need, we have demonstrated bulk extension of the optical amplification bandwidth through wideband optical parametric amplification repeaters [6] and undertaken the development of large-capacity spatial-multiplexing transmission technology using core multiplexing and mode multiplexing [7]. Our aim is to pioneer innovative extreme-high-capacity transmission technology and optical signal processing technology to make this possible and establish scalable optical network platform technology of the 1-Pbit/s class per link.

4.3 Extreme Layer 1 networking technology

We aim to generate value for users and operators through Layer 1 networking and contribute to the implementation of the APN by demonstrating specific use cases together with collaborators. To this end, we are developing elemental technology for dramatically enhancing the UX through instantaneous creation of a Layer 1 communication path to any location and elemental technology for revolutionizing operations by enabling network configuration changes to be made at will through Layer 1 switching without interrupting communications. Through these technical developments, we are exploring a variety of use cases such as demonstrating that remote e-sports matches can be conducted fairly even at the professional level by using Layer 1 communication path delay/adjustment technology [8].

5. Conclusion

This article outlined cutting-edge technologies now



WDM: wavelength division multiplexing

Fig. 4. Digital-coherent optical transmission platform technology for rollout in APN.

being pursued by NTT Network Innovation Laboratories toward IOWN/6G. Going forward, our goal is to achieve early establishment of various elemental technologies toward the deployment of IOWN/6G scheduled for 2030 by collaborating with business partners and specialists in a variety of industrial fields.

References

[1] Y. Kishiyama, S. Suyama, and S. Nagata, “Trends and Target Implementations for 5G evolution & 6G,” NTT Technical Review, Vol. 19, No. 11, pp. 18–25, Nov. 2021. https://doi.org/10.53829/ntr202111fa1
[2] Website of NTT R&D, IOWN, https://www.rd.ntt/e/iown/0002.html
[3] Specialists Center of Port and Airport Engineering (SCOPE), “Trends and Outlook for New Ocean Technologies and Unmanned Construction,” SCOPE NET, Vol. 70, Sept. 2014 (in Japanese).
[4] Press release issued by NTT, May 29, 2020 (in Japanese). https://group.ntt/jp/newsrelease/2020/05/29/200529a.html
[5] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, “6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies,” IEEE Veh. Technol. Mag., Vol. 14, No. 3, pp. 28–41, Sept. 2019.
[6] Press release issued by NTT, Jan. 28, 2021 (in Japanese). https://group.ntt/jp/newsrelease/2021/01/28/210128b.html
[7] Press release issued by NTT, “World’s First Transpacific-class Transmission Experiment Employing Mode-multiplexed Optical Signals—Key Achievement for Future High-capacity Optical Transport Network Infrastructure Supporting the Spread of 5G and IoT Services,” Mar. 9, 2020. https://group.ntt/en/newsrelease/2020/03/09/200309b.html
[8] Press release issued by NTT, “Unprecedented Optical Network Technologies to Support Esports Achieving Breathtaking Gaming UX in Remote Esports Events,” Nov. 2, 2021. https://group.ntt/en/newsrelease/2021/11/02/211102b.html

Shigeru Iwashina

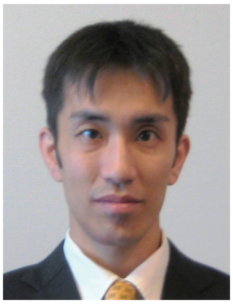
Vice President, NTT Network Innovation Laboratories.

He received a B.E. and M.E. from Tokyo Institute of Technology in 1993 and 1995. Since joining NTT DOCOMO in 1995, he has been engaged in the development of core network equipment for mobile communications systems, network architecture research, and network virtualization research. After conducting R&D on mobile communications systems ranging from 2G to 5G, he was assigned to the R&D Planning Department of NTT in July 2019 to promote R&D collaboration with partners to develop and implement NTT's IOWN initiative. He has been in his current position since July 2020.

Katsuhiro Shimano

Director, Research Planning Section, NTT Network Innovation Laboratories.

He received a B.S. from Waseda University in 1991 and M.S. from the University of Tokyo in 1993. He joined NTT in 1993, where he has been involved in R&D on optical network control and management, network virtualization, and high-performance software switches. His current interests include improving the performance of optical and wireless networks.



Koichi Takasugi

Executive Research Engineer, Director of Frontier Communication Laboratory, NTT Network Innovation Laboratories.

He received a B.E. in computer science from Tokyo Institute of Technology in 1995, M.E. from Japan Advanced Institute of Science and Technology in 1997, and Ph.D. in engineering from Waseda University in 2004. He was involved in the design and standardization of the Next-Generation Network architecture. He was also active in the AI field, such as diagnosing diabetes through machine learning. He is currently leading research on the network architecture and protocols in optical and wireless transport networks.

Kazunori Akabane

Director, Wave Propagation Laboratory, NTT Network Innovation Laboratories.

He received a B.E. and M.E. from Keio University in 1994 and 1996. He joined NTT Wireless Systems Laboratories in 1996, where he has been involved in the R&D of software-defined radio systems and IoT/machine-to-machine wireless access systems. He is currently engaged in R&D of high-speed underwater acoustic communication and terabit-class wireless transmission toward IOWN/6G.

Takashi Saida

Director, Transport Innovation Research Laboratory, NTT Network Innovation Laboratories.

He received a B.E., M.S., and Ph.D. from the University of Tokyo in 1993, 1995, and 1998. After joining NTT in 1998, he started researching advanced optical waveguide devices for communication. From 2002 to 2003, he was a visiting scholar at Ginzton Laboratory, Stanford University. From 2006 to 2008, he was a director at NTT Electronics while on leave from NTT laboratories. His current interest is in next-generation photonic transport technologies. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), the Japan Society of Applied Physics (JSAP), and the Institute of Electrical and Electronics Engineers (IEEE), and a fellow of Optica (formerly OSA).

Future Development of Digital Coherent Optical Transmission Technology

Yoshiaki Kisaka, Etsushi Yamazaki, Hideki Nishizawa, and Takashi Saida

Abstract

Digital coherent optical transmission technology—which digitally incorporates physical quantities such as amplitude, phase, and polarization and maximizes transmission performance through advanced signal processing—is the latest generation of optical transmission technology; however, the technology must be further developed to implement the Innovative Optical and Wireless Network (IOWN). The latest trends and future development of digital coherent optical transmission technology are described in this article from the perspectives of high-speed and high-capacity transmission, low-power devices, and software-based autonomous control.

Keywords: optical transmission technology, digital coherent, All-Photonics Network

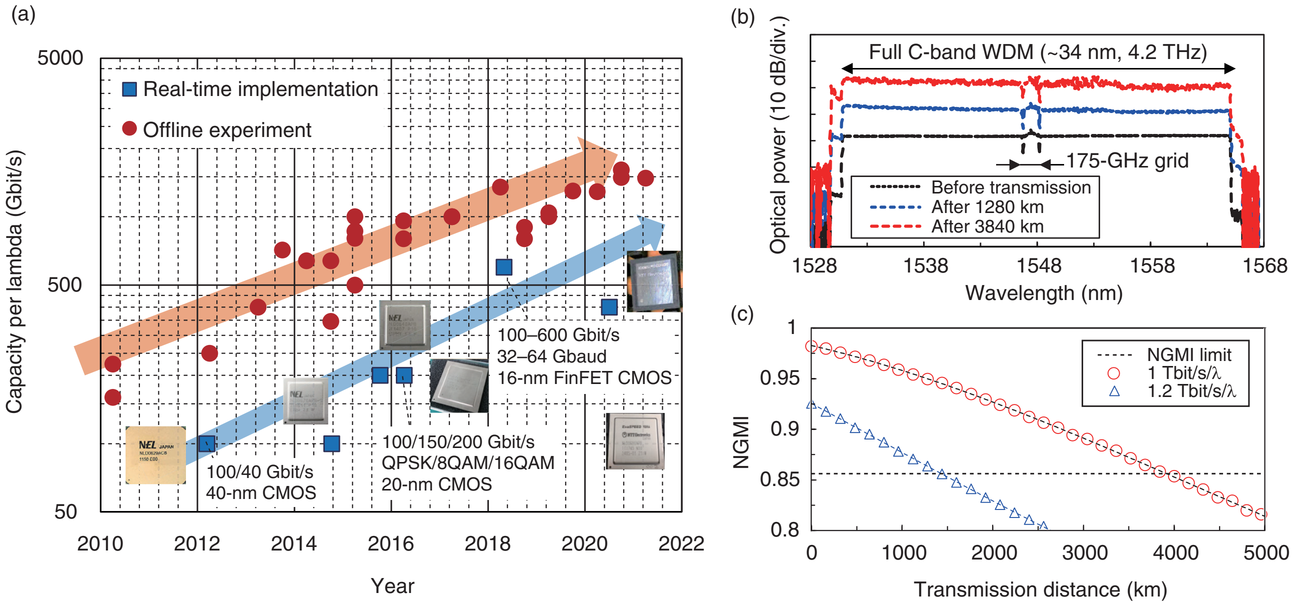
1. Introduction

Since the dawn of the Internet, Internet traffic has continued to grow exponentially in Japan and around the world as various services emerge and spread. To support this growth in traffic, optical-fiber transmission technology has continued to evolve generationally through the introduction of new technologies such as wavelength division multiplexing and optical-amplification relay. The latest generation of these technologies is digital coherent optical transmission technology, which captures all the physical quantities of light (i.e., polarization, amplitude, and phase) as digital data and compensates for distortions in optical-fiber transmission lines and optoelectronic devices through advanced signal processing [1].

To implement the Innovative Optical and Wireless Network (IOWN), which aims to integrate the real and virtual worlds into a rich society that embraces diversity, optical transmission technology must also evolve to provide even higher capacity, lower power consumption, and higher functionality. The latest trends and future development of digital coherent optical transmission technology are introduced in terms of high-speed and high-capacity transmission, compact and low-power devices, and software-based autonomous control.

2. Initiatives concerning further increase in speed and capacity

The expansion of communications within and between datacenters and the penetration of applications such as on-demand video streaming and cloud computing have further increased the demand for optical transmission with higher capacity. Regarding high-capacity networks, capacity per wavelength channel must be increased before the network can be built economically. Evolution of optical transmission capacity per wavelength channel in the cases of offline experiments and in commercial systems is shown in **Fig. 1(a)**. High-capacity transmission exceeding 1 Tbit/s per wavelength has been experimentally demonstrated. In these demonstrations, modulation speed significantly increased and modulation schemes with higher modulation levels/high-order



CMOS: complementary metal-oxide semiconductor
FinFET: fin field-effect transistor
QPSK: quadrature phase shift keying

Fig. 1. Progress of transmission capacity per wavelength and examples of transmission experiments exceeding 1 Tbit/s.

modulated signals were introduced. The modulation speed in the first generation of digital coherent optical transmission (100 Gbit/s per wavelength) was 32 Gbaud, but it has increased to 100 Gbaud and higher [2, 3]. This increased modulation speed was a result of increasing the speed of analog components such as analog-to-digital converters (ADCs) and digital-to-analog converters (DACs), optical modulators, drivers, photodiodes, and transimpedance amplifiers [4–6]. Probabilistic constellation shaping (PCS) [7], which can achieve a capacity close to the Shannon limit, has attracted attention and is used together with quadrature amplitude modulation (QAM) composed of 64 or more constellation points.

Signal distortion caused by analog components in the transmitter and receiver is a major issue in regard to increasing the capacity of optical transmission systems. As the symbol rate increases and the multi-modulation level becomes higher, signal distortions are caused by frequency-bandwidth limitation and crosstalk in the analog components and the electrical wiring in printed circuit boards and other components. These distortions are compensated for by setting optimal filter coefficients, which are estimated by calibration, for each digital equalization filter (i.e., one for transmission and one for reception). NTT Network Innovation Laboratories is developing technology to compensate for such waveform distortion with high accuracy. There are two problems with digital estimation: (i) increasing the peak-to-average power ratio and lowering signal quality [8] and (ii) tracking the time variation of waveform distortion due to operating temperature, age deterioration, and control dither. To ensure sufficient transmission performance, it is important to divide the roles of sender-side pre-equalization and receiver-side equalization. By applying our research and development (R&D) technologies to address these issues, we have commercialized an application-specific integrated circuit that achieves digital coherent transmission at a rate of up to 600 Gbit/s [9].

Regarding the development of a next-generation optical transmission system, NTT Network Innovation Laboratories has further increased modulation speed to 168 Gbaud and experimentally demonstrated transmissions of a 1-Tbit/s light signal over 3840 km and of 1.2-Tbit/s light signal over 1280 km [2]. We achieved these high modulation speeds by using two analog multiplexer (AMUX) front-end integrated modules [5], which respectively use polarization-multiplexed PCS-16QAM and 36QAM modulation schemes. This light signal is wavelength-division multiplexed (WDM) with optical frequency spacing of 175 GHz. The spectrum of the WDM signal before and after transmission is shown in **Fig. 1(b)**. The transmission line consists of pure-silica-core fiber with an 80-km optical-amplification relay section and uses both backward-pumping distributed Raman amplification and erbium-doped-fiber amplification. Each amplification achieves spectral-utilization efficiencies of 5.71 and 6.85 bit/s/Hz, respectively. Measured dependence of normal generalized mutual information (NGMI) on transmission distance for 1-Tbit/s and 1.2-Tbit/s signals is shown in **Fig. 1(c)**. As shown in the figure, NGMI values were above 0.857, namely, error-correction threshold for 21% redundancy level, for the 1-Tbit/s signal after 3840-km transmission and for the 1.2-Tbit/s signal after 1280-km transmission. This result indicates that error-free transmission is possible. Note that in this experimental demonstration, we only used the C-band, but we are conducting R&D with the aim of expanding the bandwidth to the L- and S-bands.

3. Initiatives concerning miniaturization and power reduction

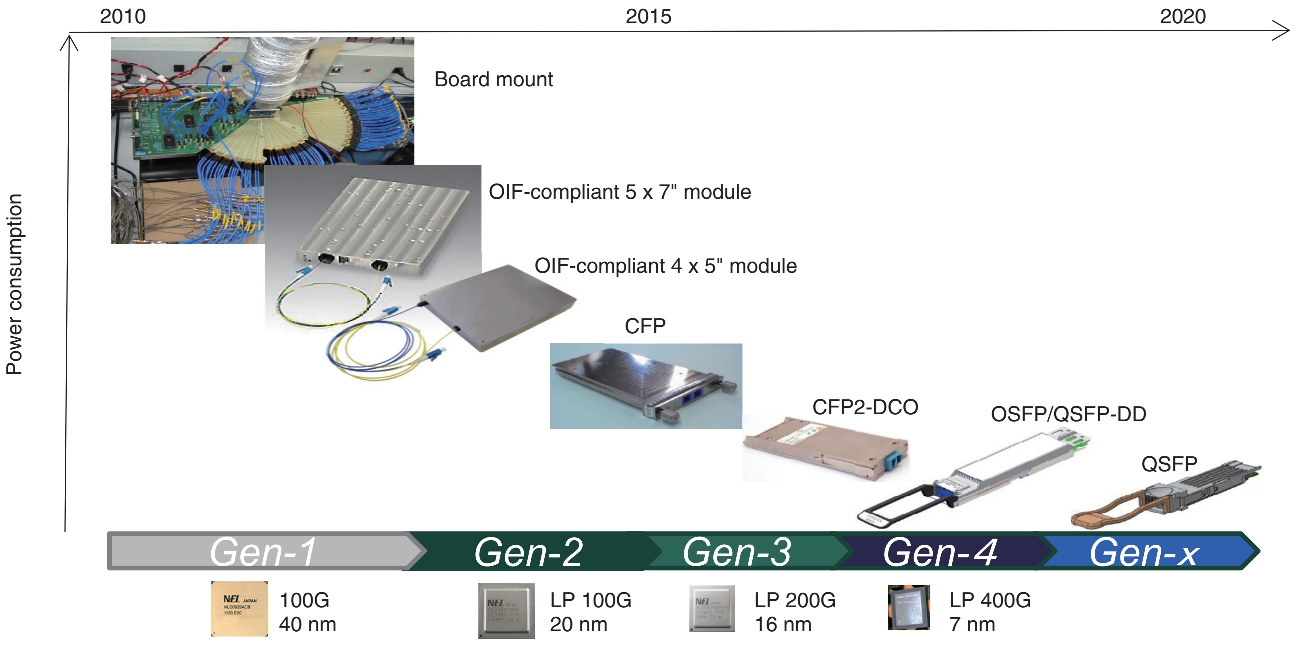
To handle ever-increasing communication traffic, optical networks must be able to continuously increase capacity. However, the installation space and power supply of optical transmission equipment are limited; accordingly, achieving higher capacity necessitates smaller, lower-power optical interfaces. Since miniaturization of devices lowers their heat-dissipation performance, low power consumption as well as miniaturization and high-density mounting of the devices that configure the optical interface are key factors in achieving miniaturization. It is therefore essential to reduce the power consumption of a digital signal processor (DSP) because it accounts for a large portion of the power consumption of optical interfaces.

Digital coherent optical transmission technology was first implemented in long-distance transmission systems configuring the backbone networks of telecommunications carriers. As the technology has matured, miniaturization, power reduction, and cost reduction have progressed, and the technology has become more widely used in short-distance transmission systems such as metro networks and datacenter-interconnect (DCI) networks. This technology is expected to be applied to access networks and networks within datacenters in the future. Low-power technologies will also become increasingly important in regard to creating future massively multi-parallel, high-capacity optical transmission systems using multicore fibers with spatial multiplexing.

The progress in miniaturization and power reduction of coherent transceivers is shown in **Fig. 2**. The DSPs developed and practically applied at NTT laboratories, i.e., installed in a coherent transceiver, are shown below the horizontal axis. Coherent transceivers are mainly used for metro and DCI networks, and their miniaturization and power-consumption reduction are rapidly progressing in a similar manner to low-power DSPs. Currently, 400-Gbit/s pluggable transceivers, such as the C form-factor pluggable 2 digital coherent optics (CFP2-DCO) and quad small form-factor pluggable double density (QSFP-DD), are in practical use.

DSPs for optical communication are required to operate at extremely high speeds and low power consumption, so the state-of-the-art complementary metal-oxide semiconductor (CMOS) process is always applied to reduce power consumption. However, this power consumption cannot be reduced sufficiently by improving the CMOS process. Accordingly, application of new signal-processing algorithms, selection of functions in accordance with application [10], and optimization of performance have progressed in regard to functions such as reduction in the digital sampling rate, wavelength-dispersion compensation, adaptive equalization, and forward-error correction (FEC). Although the application of state-of-the-art CMOS processes and innovations in signal-processing algorithms have significantly reduced the power consumption of digital circuits, the power consumption of analog circuits, such as DACs and ADCs, has not yet been significantly reduced, partly due to the effect of the higher transmission and reception signal speeds of those circuits. For that reason, the power consumption of analog circuits accounts for a larger percentage of the total power consumption in each generation of DSP, and one of the major challenges for further power reduction is to reduce the power consumption of analog circuits.

With the aim of achieving further miniaturization and lower power consumption, co-package mounting technology, in which a DSP and silicon-photonics-based optical transmitter-receiver device called coherent optical subassembly (COSA) [11] are mounted in a single package, has been researched and developed. High-density mounting of the DSP and COSA enables significant miniaturization, while shortening the high-speed analog electrical wiring



LP: low power
OIF: Optical Internetworking Forum
OSFP: octal small form-factor pluggable

Fig. 2. Miniaturization and power reduction of coherent transceivers.

between the DSP and COSA minimizes losses and keeps signal drive power low. To implement the future IOWN-based All-Photonics Network, we are conducting R&D aimed at drastically reducing power consumption by offloading functions that are currently executed by digital signal processing to optical processing. NTT Network Innovation Laboratories has demonstrated a significant improvement in optical signal-to-noise ratio (SNR) tolerance for 800-Gbit/s-class optical transmission by compensating for signal-waveform distortion caused by the bandwidth characteristics of the optical transmitter and receiver in conjunction with digital signal processing and optical processing [12]. This performance improvement makes it possible to simplify digital signal processing, thus reduce the power consumption of the DSP.

4. Initiatives concerning automated control using software

Software-defined wide-area network (SD-WAN) technology, which separates the physical network and device hardware from their control planes and manages them using software, has been commercially introduced in the market for forwarding equipment such as routers, and a system that connects remote-user sites and enables centralized control using software is now in place. In the era of full-fledged IOWN, user terminals equipped with various transmission functions, such as routers with coherent modules and white-box switches, will become increasingly popular. To connect these terminals while attaining low latency and low power consumption by reducing the number of optical-electrical conversions, using direct connection via a carrier network is being considered [13], and it is expected that software-based automatic-configuration technology for optical transmission networks will be implemented.

However, for optical transmission below layer 1, complex physical factors such as wavelength dependence of optical amplifiers and nonlinear optical effects in fiber have been considered a barrier to automated control using software as in the case of SD-WANs. To dynamically connect arbitrary user ends, a new method for quickly estimating transmission-path characteristics that determine transmission distance and capacity of optical transmission networks and real-time measurement of transmission quality using receivers are required.

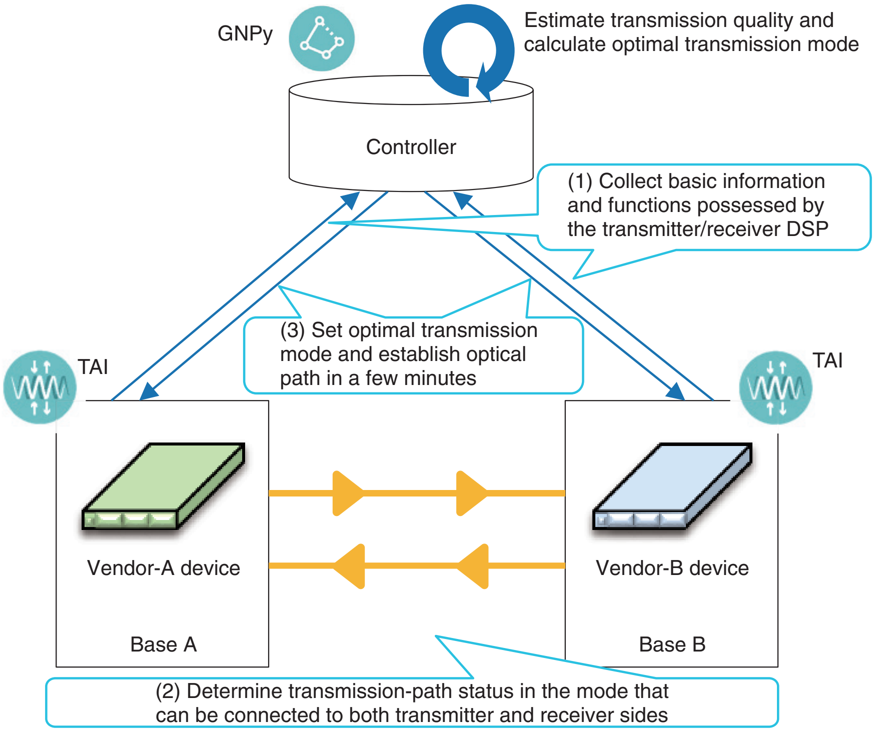


Fig. 3. Automatic optimization of optical-path transmission mode between different vendors and devices.

The Open Optical & Packet Transport project group of the Telecom Infra Project [14], which aims to define open technologies, architectures, and interfaces in optical and Internet protocol networking, is developing GNPy [15], an open-transmission-path design tool based on transmission-quality estimation using Gaussian-noise models, and a transponder abstraction interface (TAI) [16], which separates the hardware and software of a transponder. NTT Network Innovation Laboratories is contributing to openness in these areas and aims to implement a technology that automatically optimizes the transmission mode of optical paths between different vendors and devices and establishes optical-path connections in a short time by using open interfaces and tools such as a TAI and GNPy (**Fig. 3**).

Manual work has traditionally been required to connect such optical paths, and the process from service order to optical-path connection took a long time, ranging from several days to several weeks. As schemes for digital coherent optical transmission become more sophisticated and transmission modes become more diverse, the design and coordination of optical paths are becoming more complex. Our developed technology calculates the optimal transmission mode (modulation scheme and FEC type)—even in an environment in which multi-vendor transmission equipment has been deployed—by collecting such information via a TAI and estimating transmission-path conditions and transmission quality by means of GNPy and other methods. Moreover, by setting that information to the transmitter and receiver DSPs, an optical-path connection is established in a few minutes [17]. Specifically, these procedures/processes are carried out with the following sequence of steps.

(1) The basic information and functions possessed by the transmitter and receiver DSPs (configurable modulation scheme, FEC type, etc.) are collected via the TAI, and the optical-path connection is established after setting a transmission mode that can be connected to both the transmitter and receiver.

(2) Via step (1), parameters, such as bit error rate and wavelength dispersion, are collected from the receiver DSP, and the quality of the transmission channel is estimated. From these parameters, the optimal transmission mode is calculated—with appropriate margin settings—on the basis of geometric SNR estimated by GNPy.

(3) On the basis of the calculated transmission mode, parameters such as modulation method and FEC type for the transmitter and receiver DSPs are set, and an optical path between transmission devices of different vendors is opened in a few minutes.

5. Concluding remarks

To implement optical transmission technology that supports IOWN, NTT Network Innovation Laboratories will research and develop digital coherent optical transmission technology from the viewpoints of high-speed and large-capacity transmission, low-power devices, and software-based autonomous control.

References

[1] K. Kikuchi, "Fundamentals of Coherent Optical Fiber Communications," J. Lightwave Technol., Vol. 34, No. 1, pp. 157–179, 2016.

[2] M. Nakamura, F. Hamaoka, M. Nagatani, Y. Ogiso, H. Wakita, H. Yamazaki, T. Kobayashi, M. Ida, H. Nosaka, and Y. Miyamoto, "192-Gbaud Signal Generation Using Ultra-broadband Optical Frontend Module Integrated with Bandwidth Multiplexing Function," Proc. of the Optical Fiber Communication Conference and Exhibition (OFC) 2019, Th4B.4, San Diego, USA, Mar. 2019.

[3] M. Nakamura, T. Sasai, K. Saito, F. Hamaoka, T. Kobayashi, H. Yamazaki, M. Nagatani, Y. Ogiso, H. Wakita, Y. Kisaka, and Y. Miyamoto, "1.0-Tb/s/λ 3840-km and 1.2-Tb/s/λ 1280-km Transmissions with 168-GBaud PCS-QAM Signals Based on AMUX Integrated Frontend Module," Proc. of OFC 2022, San Diego, USA, Mar. 2022.

[4] M. Zhang, C. Wang, X. Chen, M. Bertrand, A. S. Ansari, S. Chandrasekhar, P. Winzer, and M. Lončar, "Ultra-high Bandwidth Integrated Lithium Niobate Modulators with Record-low Vπ," Proc. of OFC 2018, Th4A.5, San Diego, USA, Mar. 2018.

[5] Y. Ogiso, J. Ozaki, Y. Ueda, H. Wakita, M. Nagatani, H. Yamazaki, M. Nakamura, T. Kobayashi, S. Kanazawa, T. Fujii, Y. Hashizume, H. Tanobe, N. Nunoya, M. Ida, Y. Miyamoto, and M. Ishikawa, "Ultra-high Bandwidth InP IQ Modulator for Beyond 100-GBd Transmission," Proc. of OFC 2019, M2F.2, San Diego, USA, Mar. 2019.

[6] M. Nagatani, H. Wakita, H. Yamazaki, Y. Ogiso, M. Mutoh, M. Ida, F. Hamaoka, M. Nakamura, T. Kobayashi, Y. Miyamoto, and H. Nosaka, "A Beyond-1-Tb/s Coherent Optical Transmitter Front-end Based on 110-GHz-bandwidth 2:1 Analog Multiplexer in 250-nm InP DHBT," IEEE J. Solid-State Circuits, Vol. 55, No. 9, pp. 2301–2315, 2020.

[7] T. Yoshida, M. Karlsson, and E. Agrell, "Short-block-length Shaping by Simple Mark Ratio Controllers for Granular and Wide-range Spectral Efficiencies," Proc. of the 43rd European Conference on Optical Communication (ECOC 2017), Gothenburg, Sweden, Sept. 2017.

[8] T. Sasai, M. Nakamura, E. Yamazaki, A. Matsushita, S. Okamoto, K. Horikoshi, and Y. Kisaka, "Wiener-Hammerstein Model and Its Learning for Nonlinear Digital Pre-distortion of Optical Transmitters," Opt. Express, Vol. 28, No. 21, pp. 30952–30963, Oct. 2020.

[9] ExaSPEED-TERA, https://www.ntt-electronics.com/en/products/photonics/exaspeed-tera.html

[10] O. Ishida, K. Takei, and E. Yamazaki, "Power Efficient DSP Implementation for 100G-and-beyond Multi-haul Coherent Fiber-optic Communications," Proc. of OFC 2016, W3G.3, Anaheim, USA, Mar. 2016.

[11] S. Yamanaka and Y. Nasu, "Silicon Photonics Coherent Optical Subassembly for High-data-rate Signal Transmissions," Proc. of OFC 2021, Th5F.2, San Francisco, USA, June 2021.

[12] A. Matsushita, M. Nakamura, K. Horikoshi, S. Okamoto, F. Hamaoka, and Y. Kisaka, "64-GBd PDM-256QAM and 92-GBd PDM-64QAM Signal Generation Using Precise-digital-calibration aided by Optical-equalization," Proc. of OFC 2019, W4B.2, San Diego, USA, Mar. 2019.

[13] IOWN Global Forum, Technology, https://iowngf.org/technology/

[14] The Open Optical & Packet Transport group, Telecom Infra Project, https://telecominfraproject.com/oopt/

[15] GNPy, https://gnpy.readthedocs.io/en/master/

[16] TAI on GitHub, https://github.com/Telecominfraproject/oopt-tai

[17] K. Anazawa, S. Kuwabara, T. Sasai, T. Mano, T. Inoue, T. Inui, and H. Nishizawa, "Automatic Modulation-format Selection with White-box Transponders: Design and Field Trial," Proc. of the 26th Optoelectronics and Communications Conference (OECC 2021), JS2A.14, Oct. 2021.

Yoshiaki Kisaka

Senior Research Engineer, Supervisor, Transport Innovation Research Laboratory, NTT Network Innovation Laboratories.

He received a B.S. and M.S. in physics from Ritsumeikan University, Kyoto, in 1996 and 1998. In 1998, he joined NTT Optical Network Systems Laboratories, where he was engaged in R&D on high-speed optical communication systems including 40-Gbit/s/ch WDM transmission systems and the mapping/multiplexing scheme in the optical transport network (OTN). From 2007 to 2010, he joined NTT Electronics Technology Corporation, where he was engaged in the development of 40/100-Gbit/s OTN framer large-scale integration. His current research interests are in high-speed and high-capacity optical transport networks using digital coherent technology. Since 2010, he has contributed to R&D of DSP for 100-Gbit/s and beyond optical transmission. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Etsushi Yamazaki

Senior Research Engineer, Supervisor, Transport Innovation Research Laboratory, NTT Network Innovation Laboratories.

He received a B.E. and M.E. in electronic engineering from the University of Tokyo in 2000 and 2002. He is currently with NTT Network Innovation Laboratories, where he has been engaged in R&D for optical transmission systems and photonic networks. He was a visiting researcher at Stanford University from 2018 to 2019. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and IEICE. In 2012, he received the Kenjiro Sakurai Memorial Prize from the Optoelectronics Industry and Technology Development Association (OITDA) of Japan.

Hideki Nishizawa

Senior Research Engineer, Supervisor, Transport Innovation Research Laboratory, NTT Network Innovation Laboratories.

He received a B.E. and M.E. in physics from Chiba University in 1994 and 1996. In 1996, he joined NTT Network Service Systems Laboratories, where he was engaged in research on photonic switching systems. In 2001, he moved to NTT Phoenix Communications Inc. (now NTT BizLink Inc.), where he worked on the development of visual communication services. In 2007, he moved to NTT Research and Development Planning Department, where he was involved in planning R&D activities for future networks. Since 2014, he has been with NTT Network Innovation Laboratories. His current research interests include disaggregated optical transport networks that use digital coherent technology and virtualization technology. He is a member of IEICE.

Takashi Saida

Director, Transport Innovation Research Laboratory, NTT Network Innovation Laboratories.

He received a B.E., M.S., and Ph.D. from the University of Tokyo in 1993, 1995, and 1998. After joining NTT in 1998, he started researching advanced optical waveguide devices for communication. From 2002 to 2003, he was a visiting scholar at Ginzton Laboratory, Stanford University. From 2006 to 2008, he was a director of NTT Electronics while on leave from NTT laboratories. He is currently a director of NTT Network Innovation Laboratories. His current interest is in next-generation photonic transport technologies. He is a member of IEICE, the Japan Society of Applied Physics (JSAP), and IEEE and a fellow of Optica (formerly OSA).

Research and Development of Scalable Optical Transport Technologies

Yutaka Miyamoto, Kei Watanabe, and Kazuhide Nakajima

Abstract

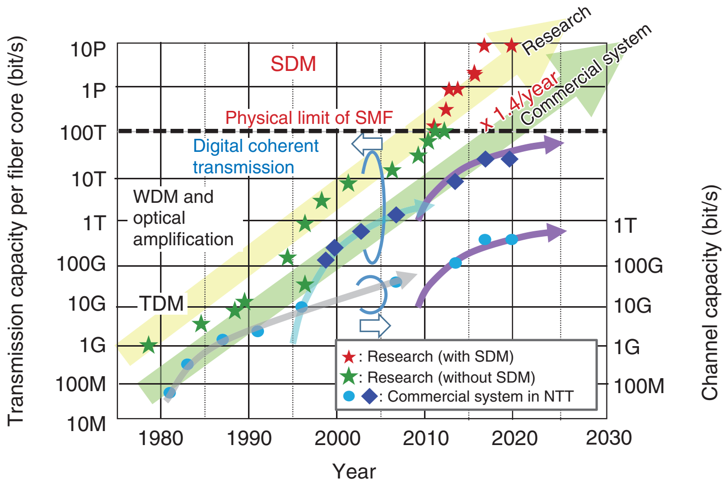
This article discusses the current state and prospects of scalable optical transport technologies that can dramatically expand the optical amplification and electrical signal processing bandwidths for achieving a Pbit/s-class long-range optical network toward IOWN (the Innovative Optical and Wireless Network) All-Photonics Network. We explain the optical parametric amplification repeater technology, which has the potential to achieve longer transmission distances while expanding the amplification bandwidth of conventional optical amplification repeaters by more than 2.5 times. We also look into the space division multiplexing optical communication technology that uses multiple-input and multiple-output signal processing and has the potential to increase transmission capacity by more than 10 times at the same cladding diameter as conventional optical fibers.

Keywords: ultra-large-capacity optical communications, optical parametric amplification repeater, mode-multiplexing optical communications

1. Introduction

The fifth-generation mobile communications (5G) system was introduced commercially in Japan in FY2020. The 5G system is forecast to evolve into an infrastructure for the Internet of things society in which all things, including autonomous driving, connect to highly reliable, low-latency, and large-scale networks. With the acceleration of the evolution of network infrastructure technologies through the development of the Innovative Optical and Wireless Network (IOWN) All-Photonics Network (APN), next-generation Beyond 5G technologies following the 5G system are expected to be widely used as network service technologies in the 2030s. The continued evolution of the network infrastructure is considered essential for flexibly supporting changes in the global industrial structure and in lifestyles brought about by the COVID-19 pandemic.

A high-capacity network infrastructure that supports the evolution of broadband services requires the continued evolution of high-capacity optical transport networks. NTT, as shown in **Fig. 1**, has been pursuing the continued development of high-capacity optical transport systems and networks based on single-mode fiber (SMF) cables, which were introduced in the 1980s. The system capacity per fiber core, along with the continuous technological innovations of various optical communication systems, has evolved at a rate of nearly 1.4 times per annum (1000 times in 20 years), wherein the system capacity is expected to exceed 1 Pbit/s by the 2030s. Digital coherent optical communication technology using digital signal processing has been commercialized as a transmission technology that takes advantage of the coherent nature of lightwave and maximizes the transmission characteristics of SMF. In 2019, an optical transport network with a capacity of 16 Tbit/s per fiber core was commercialized [1]. However, it has become clear that the physical capacity limit (capacity crunch) of SMF, a transmission medium that has supported long-distance optical transport network infrastructure, exists near the 100-Tbit/s capacity, which is approximately 10 times that of the current



SDM: space division multiplexing
TDM: time division multiplexing
WDM: wavelength division multiplexing

Fig. 1. Evolution of ultra-large-capacity optical communication system technologies.

capacity, pointing to the need for innovative technologies to achieve continued growth in system capacity.

This article describes two approaches to overcome the capacity crunch toward the IOWN APN through scalable optical transport technologies. The first is a broadband optical parametric amplification repeater technology that economically increases the SMF capacity by expanding to more than twice the optical signal bandwidth of optical amplification repeaters using conventional SMF. The second is a space division multiplexing (SDM) optical communication technology that creates multiple independent parallel communication channels multiplexed by introducing new degrees of freedom, such as multi-cores and multi-modes, in a single strand of optical fiber to achieve high-capacity communication in the Pbit/s-class capacity beyond the limit of current SMF. The following sections describe the recent progress in the individual technology elements.

2. Broadband optical parametric amplification repeater technology

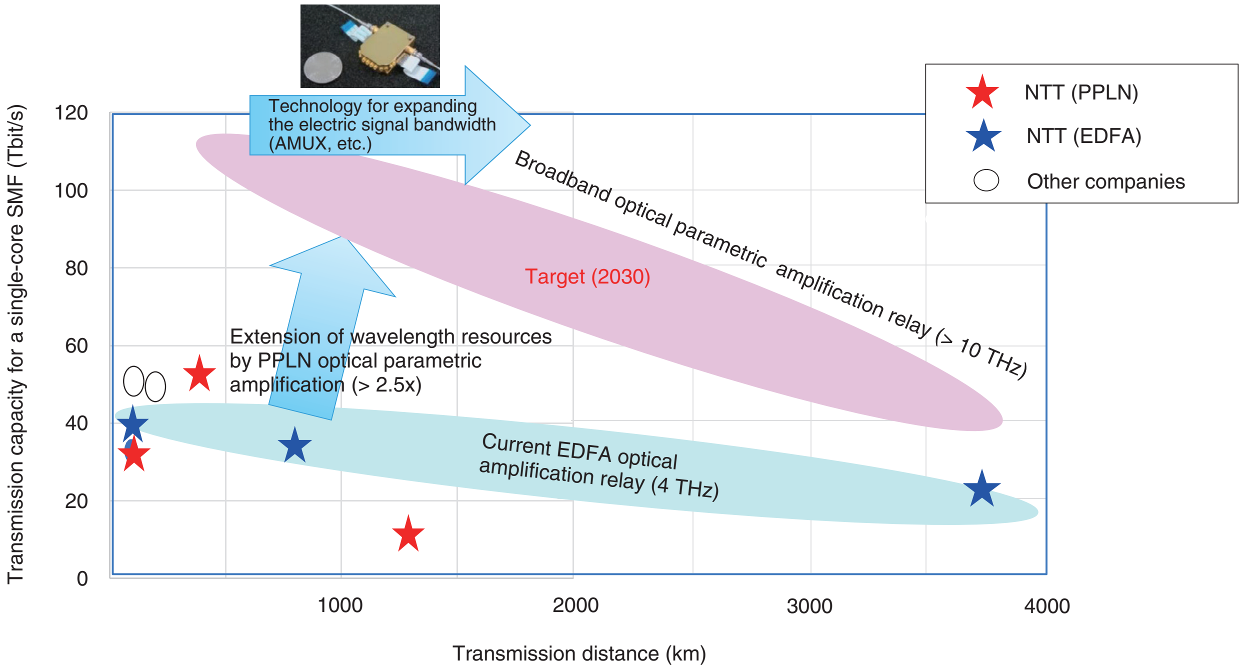
For the current optical fiber transmission systems, large-capacity transmission has been achieved by multiplexing optical signals of approximately 100 wavelengths into the optical wavelength band of around 4 THz, which is the amplification bandwidth of an erbium-doped optical fiber amplifier (EDFA), and by expanding the transmission capacity per wavelength of an optical signal using digital coherent technology[*1]. For the APN, which is part of the IOWN initiative proposed by NTT, we are aiming to build a flexible optical network that uses abundant wavelength resources. Along with increasing conventional capacity per wavelength, we are also aiming to expand the available wavelength resources (optical wavelength band). NTT has been pursuing research and development (R&D) focusing on optical parametric amplification[*2] using a periodically poled lithium niobate (PPLN)[*3] waveguide as a wide-band and low-distortion optical amplification technology

---

[*1] Digital coherent technology: A transmission method that combines massive digital signal processing and coherent detection. Coherent detection enables a high-sensitivity receiver for the amplitude and phase modulated optical signals by introducing interference between the light source placed on the receiving side and received light signal. In addition to increasing spectral efficiency through polarization multiplexing and modulation schemes using amplitude and phase of optical signals, digital coherent technology can achieve higher receiver sensitivity through high-precision distortion compensation of optical signals using digital signal processing in combination with the coherent detection.

[*2] Optical parametric amplification: Light at a specific wavelength is amplified through interaction between light of different wavelengths using the nonlinear optical effects generated in the material. High-nonlinear fibers and lithium niobate ($LiNbO_3$) crystal are known as nonlinear media.

[*3] PPLN: An artificial spontaneous-polarization crystal in which the directions of positive and negative charges in the crystal are forcibly inverted at a fixed period in a nonlinear medium, e.g., $LiNbO_3$. PPLN enables a nonlinear optical effect that is significantly higher than that of the original $LiNbO_3$ crystal.



AMUX: analog multiplexer

Fig. 2. Expansion of existing SMF transmission characteristics through broadband optical parametric amplification repeater technology and electric signal bandwidth expansion technology. All plots are experimental results with the channel rate ≥ 800 Gbit/s and optical amplification bandwidth > 4 THz.

[2]. Optical parametric amplification with a PPLN waveguide can only amplify a single polarized signal and produce phase conjugated light that is unnecessary for normal optical amplification. These factors, therefore, posed further challenges in widening the amplification bandwidth and in achieving stable response for frequent insertion and removal of optical wavelengths in response to traffic demand in the future APN with the optical amplification of polarization-multiplexed optical signals currently used in digital coherent systems.

Therefore, we proposed an amplifier configuration using multiple modular PPLN waveguides (PPLN modules) to achieve stable amplification of polarization-multiplexed optical signals and the amplification bandwidth of 10.25 THz with a gain of at least 15 dB. Using the ultra-high-speed signal-generation technology [3] with a symbol rate of over 100 Gbaud and using polarized multiplexed digital coherent signals of 800 Gbit/s per wavelength as verification signals, we confirmed low-distortion signal amplification in the gain-saturation region for both single-wavelength and wavelength-multiplexed signal inputs. We also confirmed the high-speed response to input signal switching at 1 wavelength and 41 wavelengths through emulation of the high-frequency variation in the number of wavelengths expected for the utilization of wavelength resources in the APN. We introduced the optical parametric amplifier as a wideband inline repeater and demonstrated that the optical signal bandwidth can be expanded to more than 10.25 THz, more than 2.5 times that of conventional technology, using wavelength-multiplexed signals of 800 Gbit/s per wavelength [4].

Going forward, we will pursue research on long-distance transmission performance at 100 Tbit/s, which is close to the performance limit of current optical fiber communication systems, as shown in **Fig. 2**, by combining broadband optical parametric amplification repeater technology and ultra-high-speed signal-generation technology with high symbol rate.

3. SDM optical communication technology using mode multiplexing

SDM optical communication technologies are expected to overcome the SMF capacity crunch and

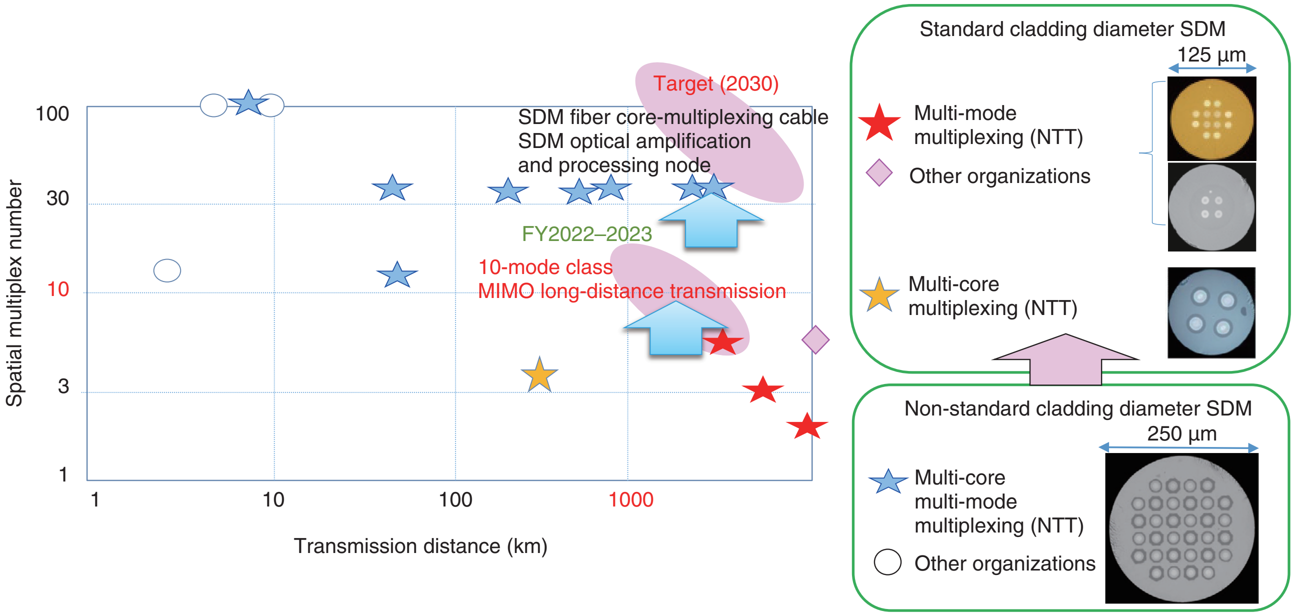


Fig. 3. Overcoming the capacity crunch using mode-multiplexed MIMO signal processing and standard cladding diameter mode-multiplexed fiber mounting technology (spatial multiplexing = 1 is the current SMF transmission system capacity).

achieve a larger capacity. In consideration of the manufacturability of SDM optical fiber cable, which serves as the novel transmission medium, should be equivalent to the standard cladding diameter (125 µm) of the currently widely used SMF. When using multiple cores and propagation modes to transmit signals using SDM optical fiber with the standard cladding diameter, especially in areas with more than four space multiplexes, strong mode couplings between each spatial mode occur, and cross talk and spatial mode dispersion (SMD) lead to significant distortion of the signal waveform. In the current digital coherent optical transmission system using SMF, polarization multiplexing is applied to transmit independent signals using two different polarizations. The dynamic waveform distortion caused by the combination of polarization rotation in the transmission path and the delay difference between the polarized waves (polarization mode dispersion) is adaptively compensated by 2 × 2 multiple-input and multiple-output (MIMO) signal processing*4 implemented in the digital signal processing circuitry in the receiver to achieve high-quality transmission. However, if this is simply expanded and applied to the mode-multiplexed SDM transmission system, the size of the MIMO signal processing circuit increases proportionately to the square of the spatial multiplex number. Furthermore, the number of digital filter taps required for MIMO signal processing must be expanded accordingly since the SMD inherent in multi-mode fiber is 10 times larger than polarization mode dispersion and accumulates proportionally to the square root of the transmission distance in multi-mode fibers with large mode coupling.

To overcome the above technical challenges, we are investigating high-capacity, long-distance optical transport technologies that actively use and control the spatial modes (**Fig. 3**). Specifically, we are aiming to establish: (1) spatial mode-control optical-fiber cabling technology with a standard cladding diameter of 125 µm suitable for fiber-optic cable installation environments and mass production, (2) mode-multiplexing MIMO processing configuration technology that takes into account the dynamic optical characteristics attributed to the cable installation properties, and (3) the fundamental technology that organically

---

*4 MIMO signal processing: A technology that transmits and receives one or more signals on a transmission path with multiple signal propagation paths (propagation modes and cores) using the same carrier frequency (wavelength). It is a widely used technology in radio communications. In optical communications, MIMO with two inputs and two outputs (2 × 2) using two orthogonal polarization modes within SMF has been commercialized using digital coherent technology as a polarization multiplexing technology.

links the spatial mode-multiplexing optical amplification repeater technology integrating (1) and (2). As an example of the results of recent studies on mode-multiplexing MIMO processing configuration technology [5], in the mode-multiplexing optical communications using six independent spatial modes, we have successfully demonstrated long-distance transmission over 6000 km by proposing an optical amplification repeater system and MIMO signal processing system that have strong compensation characteristics against transmission-loss and propagation-delay differences between different spatial modes. We have also successfully demonstrated the effectiveness of a novel implementation technology for controlling optical characteristics in mode-multiplexing transmission fiber in current terrestrial optical fiber cable structures [6]. To establish these fundamental technologies, we are accelerating R&D in cooperation with external partners, which is partially supported by the National Institute of Information and Communications Technology [7].

4. Summary

In this article, we described the current state and future prospects for broadband optical parametric amplification repeater technology and mode-multiplexed SDM optical communication technology that are being studied as scalable optical transport technologies to overcome the capacity crunch toward the IOWN APN.

References

[1] Press release issued by NTT Communications on Dec. 9, 2019 (in Japanese).
https://www.ntt.com/about-us/press-releases/news/article/2019/1209.html

[2] T. Umeki, T. Kazama, T. Kobayashi, K. Enbutsu, R. Kasahara, and Y. Miyamoto, “Low-noise Amplification and Nonlinearity Mitigation Based on Parametric Repeater Technology,” NTT Technical Review, Vol. 17, No. 5, pp. 20–26, 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201905fa3.html

[3] Press release issued by NTT, “World’s First Experimental Long-distance WDM Optical Transmission at 1 Tbps per Wavelength—Expected to Provide Large-capacity Network Communications for IoT and 5G Services,” Mar. 7, 2019.
https://group.ntt/en/newsrelease/2019/03/07/190307a.html

[4] T. Kobayashi, S. Shimizu, M. Nakamura, T. Umeki, T. Kazama, R. Kasahara, F. Hamaoka, M. Nagatani, H. Yamazaki, H. Nosaka, and Y. Miyamoto, “Wide-band Inline-amplified WDM Transmission Using PPLN-based Optical Parametric Amplifier,” IEEE J. Lightwave Technol., Vol. 39, No. 3, pp. 787–794, 2021.

[5] Press release issued by NTT, “World’s First Transpacific-class Transmission Experiment Employing Mode-multiplexed Optical Signals—Key Achievement for Future High-capacity Optical Transport Network Infrastructure Supporting the Spread of 5G and IoT Services,” Mar. 9, 2020.
https://group.ntt/en/newsrelease/2020/03/09/200309b.html

[6] Press release issued by NTT, “World’s First Attempt at Optical Transmission Property Control by Optimizing Optical Cable Structure—Milestone for Realizing Long-haul and Large-capacity Mode-multiplexing Optical Transmission Links,” Mar. 9, 2020.
https://group.ntt/en/newsrelease/2020/03/09/200309a.html

[7] Y. Miyamoto, K. Nakajima, T. Hasegawa, E. L. T. de Gabory, R. Sugizaki, and R. Nasase, “R&D of Spatial-mode-controllable Optical Transmission System,” IEICE General Conference, Planning Symposium, BI-3-1, Mar. 2022 (in Japanese).

Yutaka Miyamoto

NTT Fellow, Director, Innovative Photonic Network Research Center, NTT Network Innovation Laboratories.

He received a B.E. and M.E. in electrical engineering from Waseda University, Tokyo, in 1986 and 1988 and Dr.Eng. in electrical engineering from the University of Tokyo in 2016. He joined NTT Transmission Systems Laboratories in 1988, where he engaged in R&D on high-speed optical communications systems including the 10-Gbit/s first terrestrial optical transmission system (FA-10G) using EDFA inline repeaters. He was with NTT Electronics Technology Corporation between 1995 and 1997, where he engaged in the planning and product development of a high-speed optical module at the data rate of 10 Gbit/s and beyond. Since 1997, he has been with NTT Network Innovation Labs, where he has researched and developed optical transport technologies based on 40/100/400-Gbit/s channel and beyond. He is currently investigating and promoting the future scalable optical transport network with Pbit/s-class capacity. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and a fellow of the Institute of Electronics, Information and Communication Engineers (IEICE).

Kei Watanabe

Senior Research Engineer, Heterogeneous Materials and Devices Research Group, NTT Device Technology Laboratories.

He received a B.E. in electrical and electronics engineering and M.E. and Dr.Eng. in mathematical and material science from Kobe University, Hyogo, in 1998, 2000, and 2003. In 2003, he joined NTT Photonics Laboratories, where he engaged in research on silica-based optical waveguides. From 2008 to 2009, he was a visiting researcher at Ghent University, Belgium, where he engaged in research on silicon photonics. After returning to NTT, he has been engaged in research on indium phosphide high-speed modulators and nonlinear optical devices based on PPLN waveguides. He is a member of the Japan Society of Applied Physics (JSAP), IEICE, and the IEEE Photonics Society.

Kazuhide Nakajima

Senior Distinguished Researcher, NTT Access Network Service Systems Laboratories.

He received an M.S. and Ph.D. in electrical engineering from Nihon University, Chiba, in 1994 and 2005. In 1994, he joined NTT Access Network Service Systems Laboratories, where he has been engaged in research on optical-fiber design and related measurement techniques. He has been acting as a rapporteur of Question 5 of the ITU-T (International Telecommunication Union - Telecommunication Standardization Sector) Study Group 15 since 2009. He is a member of IEEE, Optica (formerly OSA), IEICE, and JSAP.

R&D Activities of Core Wireless Technologies toward 6G Radio Access

Jun Mashino, Yosuke Fujino, and Riichi Kudo

Abstract

Technical research for the 6th-generation mobile communication system (6G) expected to be implemented in the 2030s has been advanced to achieve extreme high data rate/capacity, extreme low latency, extreme high reliability, and extreme coverage extension for non-terrestrial areas. In this article, we introduce the following core wireless technologies developed by NTT Network Innovation Laboratories toward 6G, i.e., orbital angular momentum multiple-input multiple-output (OAM-MIMO) multiplexing transmission, underwater acoustic communication, and wireless-link-quality prediction.

Keywords: OAM multiplexing, underwater acoustic communication, wireless-link-quality prediction

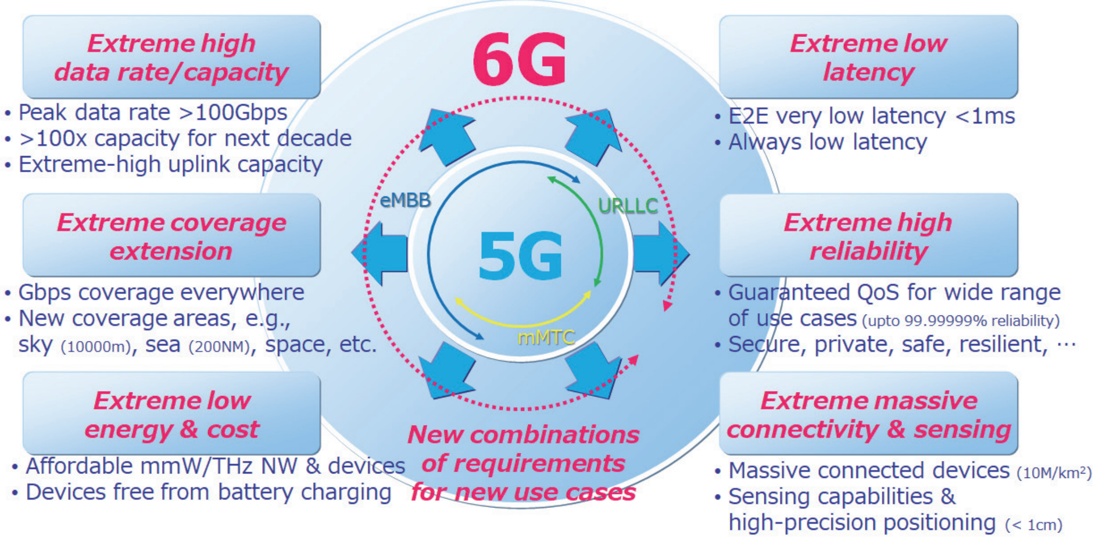
1. Introduction

Mobile communication systems, which are now essential for our daily lives, have evolved once every ten years. The 5th-generation mobile communication system (5G) service has been provided since 2020 in Japan, and much research for 6G, expected to be implemented in the 2030s, has been conducted around the world. The vision with mobile communication systems has been providing mobile multimedia functions including displaying websites, movie content, and reinforcement of application software in the 3G/4G era. The vision in the 5G/6G era will be expected to take on a role as a society infrastructure that can address various social issues and support diverse industries.

Use cases, target performance, and technology candidates expected for 5G evolution and 6G are described in NTT DOCOMO's 6G White Paper [1], including the concept of 6G combined with innovative networking and information technologies for the Innovative Optical and Wireless Network (IOWN) led by the NTT Group. Technical development in 6G addresses not only basic performance enhancement (e.g., extreme high data rate/capacity, extreme low latency, extreme high reliability, and extreme massive connectivity) but also new challenges including extreme coverage extension to the sky, sea, and space (**Fig. 1**). In this article, we introduce the following three core wireless technologies developed by NTT Network Innovation Laboratories toward 6G, i.e., orbital angular momentum multiple-input multiple-output (OAM-MIMO) multiplexing transmission, underwater acoustic communication, and wireless-link-quality prediction.

2. OAM-MIMO multiplexing transmission for terabit-class wireless communication

The benefit of 5G is high-speed transmission by introducing wide bandwidth in the high-frequency band called millimeter waves, which is being used in mobile communication systems for the first time. The final target in 5G radio access is 20-Gbit/s transmission in future evolution, and 6G radio access is expected to reach 100-Gbit/s transmission by exploring a higher-frequency band known as the sub-THz band to obtain more radio resources. However, a high density of base-station antennas has to be deployed compared with current systems due to the problems of straightness of radio waves and coverage holes due to blocking in such a higher-frequency band. Therefore, a flexible network and easy installability are required for xHaul* connecting between base-station equipment. If wireless high-capacity xHaul is achieved, it will be possible to provide a highly dense



Source: DOCOMO 6G White Paper [1]

E2E: end to end
eMBB: enhanced mobile broadband
mMTC: massive machine-type communications
QoS: quality of service
URLLC: ultra-reliable and low latency communications

Fig. 1. Requirements for 6G wireless technology.

deployment solution of base-station equipment in areas where there is difficulty in optical wiring or temporary expansion of base-station equipment. However, wireless xHaul for 6G requires an extremely high capacity of over 1 Tbit/s (terabit-class) transmission in consideration of radio access evolution described above, function split between base-station equipment, and accommodation of multiple types of base-station equipment (**Fig. 2**).

NTT Network Innovation Laboratories has conducted advanced research on OAM multiplexing technology. OAM is a physical property of electromagnetic waves characterized by a helical phase front in the propagation direction. Since these characteristics can be used to create multiple orthogonal channels, wireless communication using OAM can increase radio-spectrum efficiency. Given an electromagnetic wave having this OAM property, the trace of the same phase takes on a helical shape in the propagation direction. Different data signals can be transmitted simultaneously by transmitting different signals using radio waves having different OAM modes. Since radio waves having this OAM property cannot be received without a receiver having the same number of phase rotations at the time of transmission, multiple radio waves having different OAM modes can eventually be separated without mutual interference.

OAM-MIMO multiplexing transmission, originally proposed by NTT, uses coaxial uniform circular array (UCA) antennas for OAM transmission and reception in consideration of practical implementation. We combined OAM and existing MIMO technology by treating UCAs with different radii as sub-array antennas for MIMO transmission so that an extreme amount of spatial multiplexing can be achieved while preventing the use of higher-order OAM modes, which is inappropriate for long-distance transmission (**Fig. 3**). Load reduction for baseband digital signal processing is one of the advantages of OAM-MIMO since signal processing can be separated into analog passive circuits for OAM and digital signal processing for MIMO. This contributes to low-energy consumption and achieving transmission with much wider bandwidth.

An indoor experimental trial was conducted in the 28-GHz band to challenge the limit of spatial multiplexing. The results indicate that OAM-MIMO can multiplex 21 spatial streams in combination with vertical/horizontal polarization and achieve high-capacity transmission of 201.5 Gbit/s [2] (**Fig. 4**). An outdoor experimental trial was also conducted in the 40-GHz band to prove the feasibility of long-distance

---

[*] xHaul: Unified transport network infrastructure connecting between base-station equipment, e.g., fronthaul, midhaul, and backhaul.

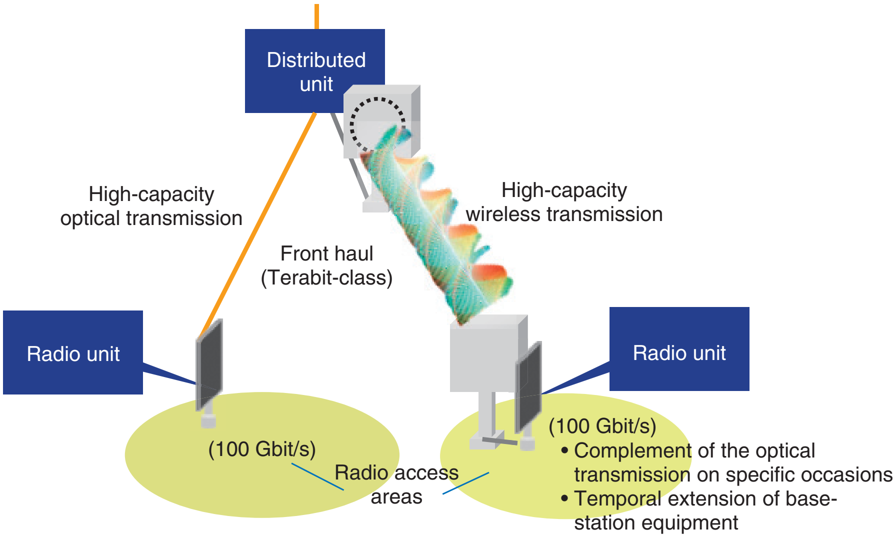


Fig. 2. An example of xHaul for base-station equipment.

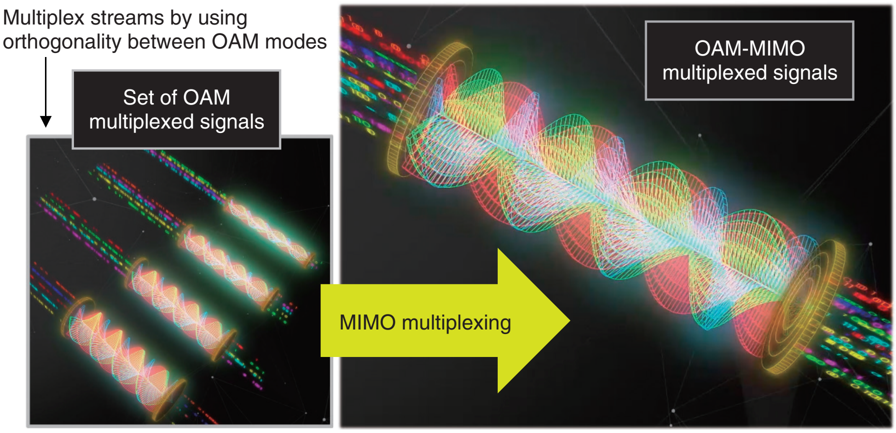


Fig. 3. Concept of OAM-MIMO multiplexing transmission.

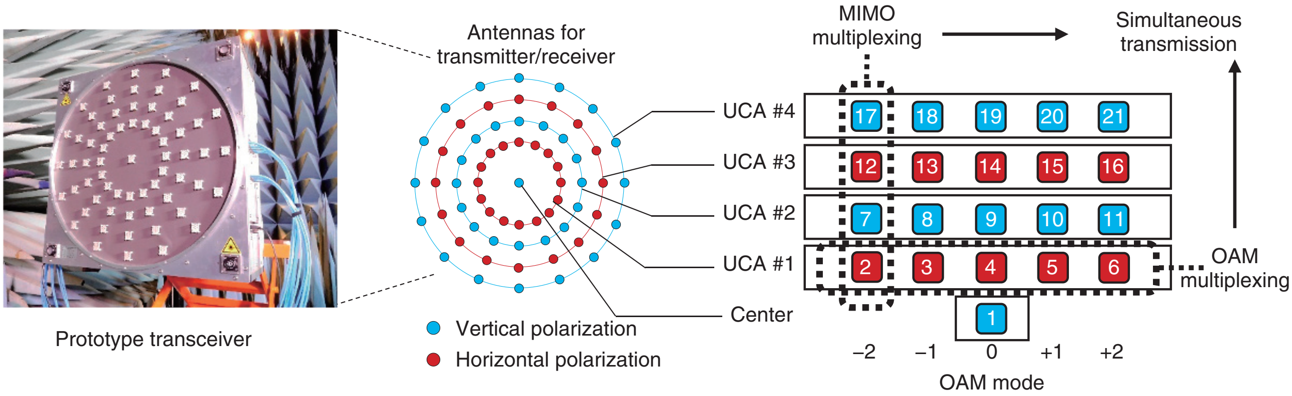


Fig. 4. Transceiver antennas and assignment of 21 spatial streams.

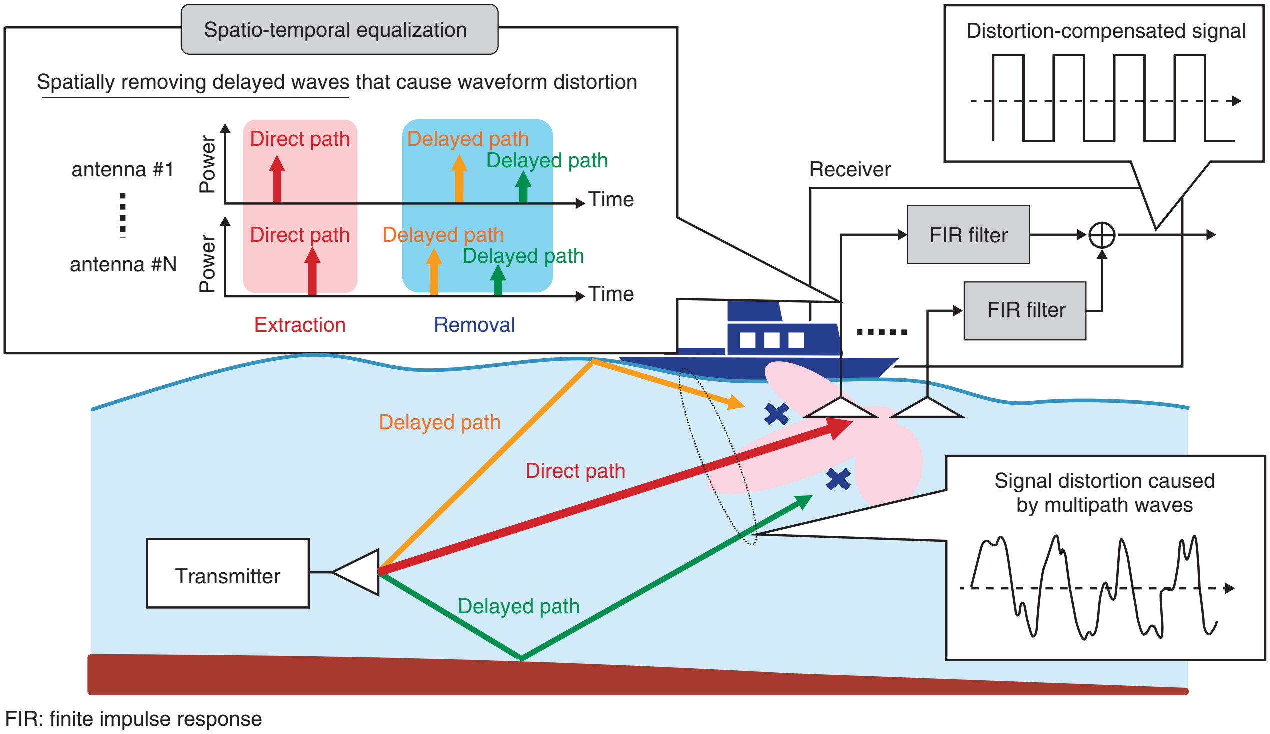


Fig. 5. Concept of spatio-temporal equalization.

transmission. It was successful in 117-Gbit/s transmission for 200 m [3]. We are now pushing forward with our research to achieve terabit-class wireless communication required for xHaul in 6G by expanding the bandwidth up to 10 GHz and combining it with 20–40 spatial streams in advanced OAM-MIMO technologies.

3. Underwater acoustic communication to extend coverage into the ocean

The development of offshore resources, such as oil and gas fields, has been growing, and remotely operated vehicles (ROVs) are used for offshore development. Unmanned remote construction using ROVs is also being researched to improve worker safety and operational efficiency in port and harbor construction. Since the Mbit/s-class underwater communication capable of transmitting high-definition video has not yet been commercialized, these ROVs are connected to support vessels on the sea by long wired cables. The main drawback of wire-controlled ROVs is their high operational cost due to large support vessels and dedicated operators to hoist the long and heavy cables. Therefore, Mbit/s-class high-speed underwater communication that enables wireless remotely controlled ROVs is highly desired. Underwater communication using various media such as radio waves, sound waves, and light have been studied. We have focused on sound waves, which are suitable for stable long-distance communication, and are working to improve the speed of underwater acoustic communication.

Not only in underwater acoustic communication but in any wireless communication, waveform distortion occurs as a result of combining many waves from different paths. Therefore, wireless communication requires equalization to compensate for this waveform distortion. Due to the extremely slow propagation speed of sound waves compared with radio waves, waveform distortion drastically changes in underwater acoustic communication, which makes equalization more complicated. The conventional equalization technique, which compensates for the inverse response of waveform distortion, cannot track fast fluctuations in waveform distortion of underwater acoustic communication because the inverse response cannot be estimated in time. As a result, the communication speed of conventional underwater acoustic communication has been limited to several tens of kbit/s [4].

To overcome the fast fluctuations in waveform distortion, we proposed an equalization technique called spatio-temporal equalization [5] (**Fig. 5**). This technique

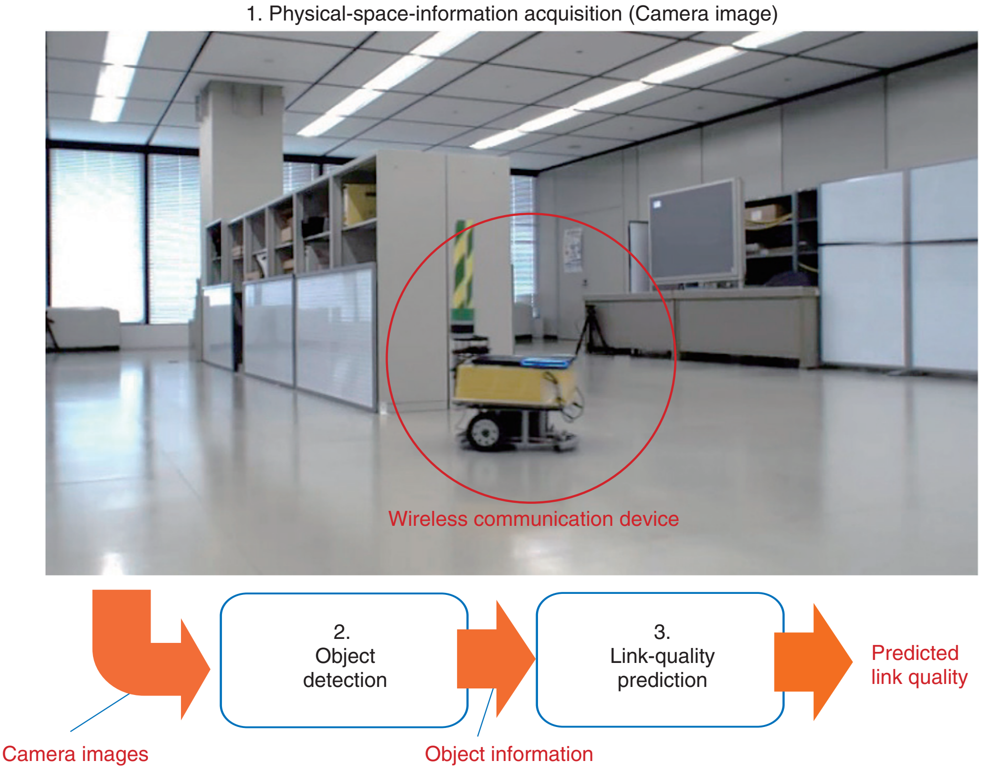


Fig. 6. Link-quality prediction using camera images.

achieves equalization by spatially removing the delayed waves that cause waveform distortion using adaptive beamforming. Spatio-temporal equalization does not require the estimation of the inverse response of waveform distortion, enabling Mbit/s-class high-speed acoustic communication.

We conducted transmission experiments in the ocean to verify the effectiveness of this technology and succeeded in transmitting 5.12 Mbit/s at a distance of 18 m and 1.2 Mbit/s at a distance of 60 m [6]. We are currently working on achieving more than 1 Mbit/s at a distance of 300 m by combining bandwidth-division transmission technology, which uses multiple transmitters, with different resonance frequencies to transmit broadband signals.

4. Wireless-link-quality-prediction technologies using physical space information

In the 6G era, the use of physical space information will be more familiar because of various activities for Society 5.0 [7], which was proposed as a future society that Japan should aspire to. The huge amount of physical space information will be stored in cyber space and become more accessible from everywhere. NTT Network Innovation Laboratories started to research the physical space information use to promote the evolution to wireless communication systems [8, 9]. The throughput and capacity in wireless communication systems have been basically enhanced using wireless system information and settings. The requirements for wireless communication systems are diversified and advanced because of emerging services. This raises the expectation of highly reliable wireless accesses. Physical space information use is one of the keys to evolve wireless communication systems since the long-term movement of communication devices and surrounding objects can be extracted from the physical space information. Long-term prediction enables proactive actions with sufficient time to prepare the advanced controls and develop the roles of wireless communication systems.

Throughput prediction using camera images is shown in **Fig. 6** as an example of our work. Cameras first obtain images including communication device. A device is then detected from the obtained images by using object-detection algorithms with a pre-learned model. Finally, the wireless-link-quality-prediction model predicts link quality by using the detected object information. The exteriors of communication devices are learned using object-detection algorithms that are based on deep learning, and the detected position and movements of communication devices on the image are used to predict future link quality. Since a large number of cameras and sensors are expected to be connected to the IOWN All-Photonics Network, sustainable system design is also being researched to enable the addition and deletion of cameras and sensors for prediction systems. The artificial-intelligence-driven experimental rooms in which multimodal features including physical space information and wireless system parameters are automatically measured are being developed to evaluate the performances of various machine learning technologies using big data measured in an experimental environment. The accurate prediction in the short and long term enables proactive controls, such as switching to better wireless connections, to change the application data rate or control the movement of communication devices and surrounding objects by detecting the link disconnection beforehand. We believe that physical-space-information-based wireless-link-quality prediction will extend the use cases of reliable wireless communication.

References

[1] NTT DOCOMO, "DOCOMO 6G White Paper," https://www.docomo.ne.jp/english/corporate/technology/whitepaper_6g/

[2] Y. Yagi, H. Sasaki, T. Yamada, and D. Lee, "200 Gbit/s Wireless Transmission Using Dual-polarized OAM-MIMO Multiplexing on 28 GHz Band," Proc. of IEEE Globecom Workshops on HCPtP, Waikoloa, USA, Dec. 2019.

[3] Y. Yagi, H. Sasaki, T. Semoto, T. Kageyama, T. Yamada, J. Mashino, and D. Lee, "Field Experiment of 117 Gbit/s Wireless Transmission Using OAM Multiplexing at a Distance of 200 m on a 40 GHz Band," Proc. of IEEE International Conference on Communications Workshops, Montreal, Canada, June 2021.

[4] S. Sendra, J. Lloret, J. M. Jimenez, and L. Parra, "Underwater Acoustic Modems," IEEE Sens. J., Vol. 16, No. 11, pp. 4063–4071, June 2016.

[5] Y. Fujino and H. Fukumoto, "Spatio-temporal Equalization Technology for High-speed Underwater Acoustic Communications," IEICE Commun. Mag., Vol. 60, pp. 284–297, Mar. 2022 (in Japanese).

[6] H. Fukumoto, Y. Fujino, M. Nakano, K. Sakamoto, and T. Tsubaki, "Field Experiments Demonstrating Mbps-class Underwater Acoustic Communication with Spatio-temporal Equalization," Proc. of IEEE Global OCEANS 2020, Biloxi, USA, Oct. 2020.

[7] Cabinet Office, Government of Japan, "Society 5.0," https://www8.cao.go.jp/cstp/english/society5_0/index.html

[8] R. Kudo, M. Cochrane, K. Takahashi, T. Inoue, and K. Mizuno, "Experimental Validation of Link Quality Prediction Using Exact Self-status of Mobility Robots in Wireless LAN Systems," IEICE Trans. Commun., Vol. E103-B, No. 12, pp. 1385–1393, July 2020.

[9] R. Kudo, K. Takahashi, T. Inoue, and K. Mizuno, "Using Vision-based Object Detection for Link Quality Prediction in 5.6-GHz Channel," EURASIP J. Wireless Com. Network., 207, Oct. 2020.

Jun Mashino

Senior Research Engineer, NTT Network Innovation Laboratories.

He received a B.E. in electrical and electronic engineering and M.E. in communications and computer engineering from Kyoto University in 2003 and 2005. He joined NTT Access Network Service Systems Laboratories in 2005, where he has been engaged in the research and development of intelligent interference compensation technologies and signal processing for future wireless communication systems. From 2016 to 2019, he engineered uplink massive MIMO technology for 5G as a senior research engineer at NTT DOCOMO 5G Laboratory and led proof of concept projects for social implement of 5G. Since 2019, he has been a senior research engineer at NTT Network Innovation Laboratories. His current research interests are in high-capacity wireless transmission technologies including OAM-MIMO toward 6G.

Yosuke Fujino

Senior Research Engineer, NTT Network Innovation Laboratories.

He received a B.E. and M.E. in electrical and electronic engineering from Shizuoka University in 2002 and 2004. He joined NTT Network Innovation Laboratories in 2004 and engaged in research on physical layer signal design, transceiver architecture, and signal processing for Internet of Things wireless systems. The current focus of his research is wireless communication technology in non-terrestrial areas such as underwater, underground, and space.

Riichi Kudo

Senior Research Engineer, Supervisor, NTT Network Innovation Laboratories.

He received a B.S. and M.S. in geophysics from Tohoku University, Miyagi, in 2001 and 2003 and Ph.D. in informatics from Kyoto University in 2010. In 2003, he joined NTT Network Innovation Laboratories. He was a visiting fellow at the Center for Communications Research (CCR), Bristol University, UK, from 2012 to 2013, and worked for NTT DOCOMO from 2015 to 2018. He is now working for NTT Network Innovation Laboratories.

Overview of Technical Development and Verification in the Connected-vehicle Field

Atsushi Koizumi and Shin Mitsuhashi

Abstract

As technological innovations, such as connectivity, automation, sharing, and electrification, bring major changes to the automotive industry, expectations for and the importance of information and communication technology (ICT) are increasing rapidly in the industry. The NTT Group and Toyota Motor Corporation are collaborating on the research and development of an ICT platform for connected vehicles. They conducted joint field trials from 2018 to 2020 and established the basic technology through various use cases and verification of the platform. In the Feature Articles in this issue, the NTT Group operating companies and NTT laboratories that are participating in the collaboration present the details of the field trials, the results, technologies applied, value provided, and future issues.

Keywords: connected vehicle, IoT, big data

1. Overview of connected vehicles

A connected vehicle, which is a mobility device equipped with communication capability, consists of various components, including a vehicle, communication networks (wired and wireless), and cloud computing (**Fig. 1**). To implement these components, it is necessary to develop, use, and combine a wide range of technologies, including networks for exchanging data, edge computing for data processing in the proximity of connected vehicles, a platform for storing and processing collected data, platform for analyzing and using data, and software updates.

As the potential for using big data expands, the volume of data to handle is expected to increase dramatically. Therefore, information and communication technology (ICT) platforms, such as networks and datacenters, that receive data from connected vehicles are growing in importance.

2. Potential of vehicle big data

Connected vehicles are already in use, and the market for them will continue to expand. If the large amount of data held by connected vehicles can be processed quickly and at low cost, it will become possible to use information that cannot be captured with current sensors and provide faster and more reliable services. This will in turn not only improve convenience and efficiency but also enable people to drive more safely and securely, alleviate or eliminate traffic congestion, and shorten travel time, which will contribute to achieving carbon neutrality (**Fig. 2**).

3. Purpose and areas of joint research and development

In March 2017, the NTT Group and Toyota Motor Corporation agreed to collaborate in developing, verifying, and standardizing technologies needed in the connected-vehicle field by combining Toyota's vehicle-related technologies and NTT Group companies' ICT-related technologies.

By sharing the technologies and expertise of each company and using big data obtained from vehicles, the NTT Group and Toyota will work together to research and develop the technologies needed to solve problems facing society, such as traffic accidents

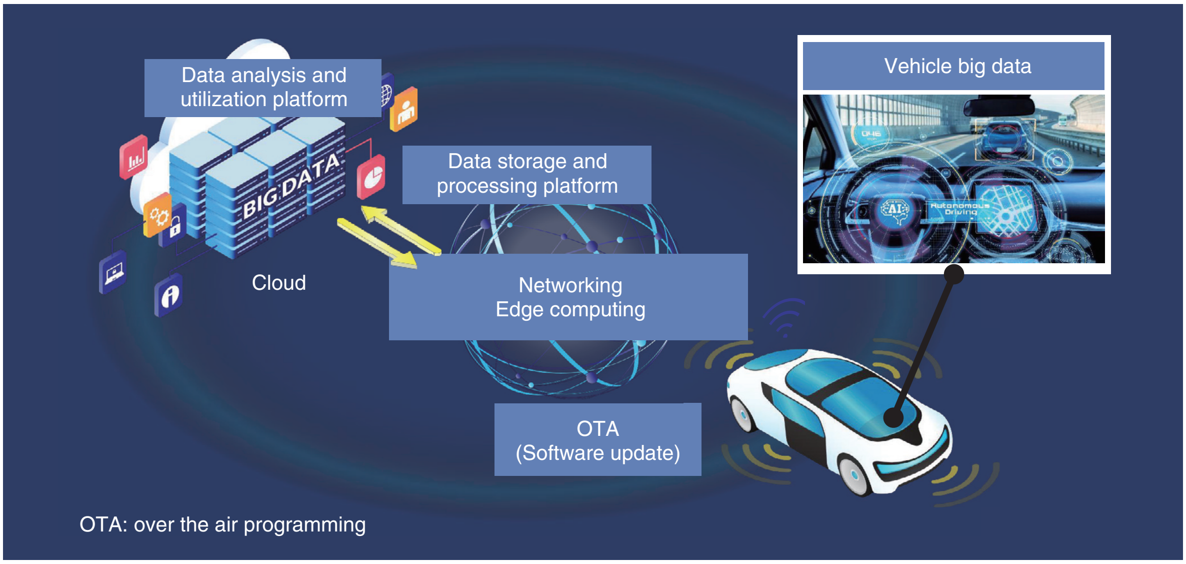


Fig. 1. Overview of a connected vehicle.

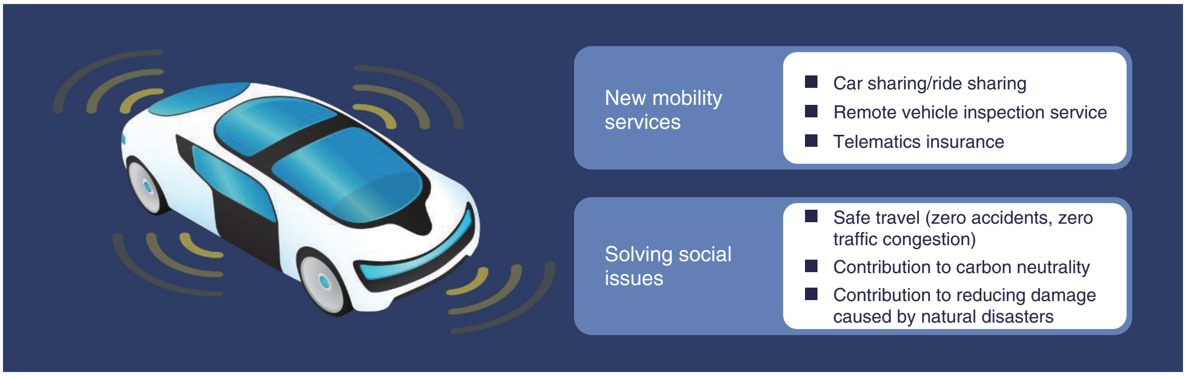


Fig. 2. Value provided by vehicle data.

and congestion, and provide new mobility services to customers. The collaboration is aimed at creating a sustainable smart mobility society from a global perspective.

The collaboration activities can be broadly divided into the following three technical areas:

(1) Data collection/storage/analysis platform: Establish a mechanism for storing and analyzing data sent from millions or tens of millions of vehicles.

(2) Internet of Things (IoT) network and datacenter: Establish an optimal arrangement of networks and datacenters for collecting data from vehicles around the world.

(3) Next-generation communication technology: Verify technology for vehicles to use the 5th-generation mobile communication system (5G) and verify the applicability of edge computing.

In December 2018, we set the goal of processing a large volume of data from within and outside vehicles to reproduce the physical space in the real world in real time (within seconds) and with a precision of tens of centimeters. We carried out field trials that covered the end-to-end mobility system including vehicles, networks, and datacenters.

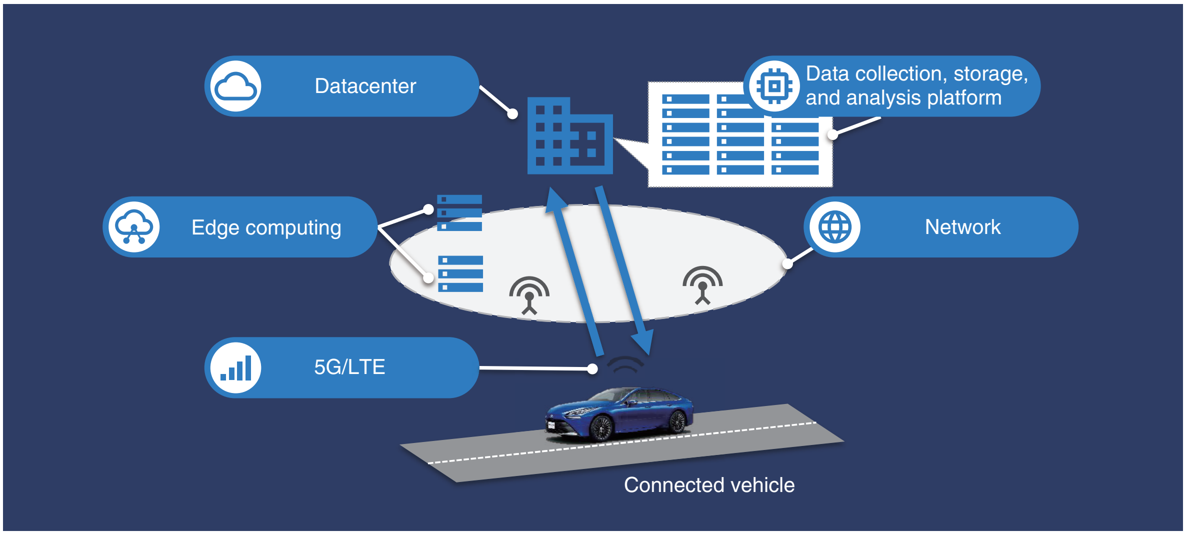


Fig. 3. Overview of the ICT platform for connected vehicles.

4. ICT platform for connected vehicles

The ICT platform for connected vehicles connects vehicles equipped with communication capability with edge computing nodes and the cloud through LTE (Long-Term Evolution), 5G, and IoT networks. It also uses computing resources to collect, store, and analyze data that are held or collected by vehicles (**Fig. 3**).

Data from connected vehicles are uploaded to datacenters through the mobile network and edge nodes. The datacenter then collects, stores, and analyzes the data and returns the analysis results to vehicles through the network as necessary. Although this data flow is not much different from that of smartphones and small IoT devices, a mechanism for real-time, large-scale, and accurate processing of such data will become one of the important infrastructures for supporting people's daily lives and society at large.

5. Method of conducting joint research and development

The joint research and development was conducted in two alternating activities: technical studies by a working group and verification of its study results using actual vehicles (**Fig. 4**).

In the working group weekly meetings, engineers from the two parties discussed multiple themes in parallel. The working group's discussion results were verified using a testbed that involved more than a hundred physical servers, 5G, other communication links, and actual vehicles. The verification results were then fed back to the working group. This cycle was then repeated over a short period. The working group members were proud of the fact that the two parties grew together by bringing together their technologies and expertise and by repeating the above cycle, therefore quickly solved various technical issues.

6. Activities in the collaboration on connected vehicles

The field trials of the ICT platform for connected vehicles were conducted from FY2018 to FY2020. The following two articles in this issue give an overview and the details of the field trials from the perspectives of NTT DATA and NTT Communications, which played a central role in the trials. These articles also introduce the technical results and achievements obtained and future challenges identified in the trials.

(1) Activities and results of field trials—reference architecture for a connected-vehicle platform: The overall architecture and verification of the platform through use cases and verification results are presented [2].

(2) Activities and results of field trials—network edge computing platform: An overview, the features, and methods of the platform are presented [3].

The ICT platform for connected vehicles involves a

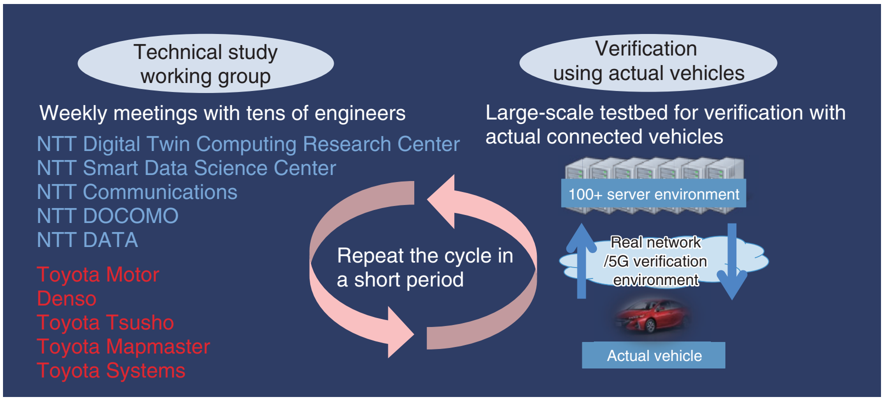


Fig. 4. Method of conducting the joint research and development.

variety of technical challenges. The following five articles in this issue focus on representative technologies that have been studied and developed to solve these challenges and provide overviews of these technologies and value they provide, including verification through use cases and field trials, and identify remaining challenges.

(3) Real-time spatiotemporal data-management technology (Axispot™): This technology stores, searches, and analyzes large volumes of data on dynamic objects [4].

(4) Selective vehicle-data-collection algorithm: This technology determines data-collection priorities on the basis of meta-information, such as vehicle positions and time, observation range, and the effects of shielding by surrounding vehicles [5].

(5) Vertically distributed computing technology: This technology effectively uses limited server resources by dynamically changing the response-processing server on the basis of vehicle status [6].

(6) Lane-specific traffic-jam-detection technology: This technology detects lane-specific traffic jams by collecting and analyzing dashcam video and driving data to achieve optimal lane navigation [7].

(7) Technology for calculating suddenness index for aggregated values: To reduce the amount of processing on the ICT platform, this method calculates an index for periodic and sudden changes in aggregated values collected from vehicles on the basis of the degree of deviation from the ordinary state [8].

7. Results and future outlook

"The Technical Document on the ICT Platform for Connected Vehicles" was compiled and published in November 2021 to report on the results and challenges of the three-year field trials [9]. We are hoping that the results of our activities will be widely used not only by those in the ICT and mobility industries but also by those in other industries.

The NTT Group and Toyota will continue to improve the speed, efficiency, and sophistication of the ICT platform for connected vehicles in preparation for the further spread of such vehicles. We will also continue to develop technologies that will contribute to solving problems facing society, such as traffic accidents and congestion, by effectively using big data collected from vehicles and by creating value from such data.

We will also use the obtained technical results to plan social implementation of the ICT platform for connected vehicles and the deployment of the verified technologies in smart cities. We will develop technologies for providing high-speed, high-capacity communications and vast computing resources beyond the limits of the conventional platform and collaborate with various companies, organizations, and service providers to create and provide new mobility services, thereby contributing to achieving carbon neutrality and creating a sustainable smart mobility society that brings safety and security to people.

References

[1] Press release issued by Toyota and NTT, "Toyota and NTT Agree to

Collaborate on ICT Platform R&D for Connected Cars Collaboration Seeks to Establish Technologies for the Creation of a Smart Mobility Society," Mar. 27, 2017.
https://group.ntt/en/newsrelease/2017/03/27/170327a.html

[2] Y. Chiba, "Activities and Results of Field Trials—Reference Architecture for a Connected-vehicle Platform," NTT Technical Review, Vol. 20, No. 7, pp. 42–47, July 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202207fa6.html

[3] R. Noji, T. Kamei, K. Funabiki, D. Murata, and Y. Kanemaru, "Activities and Results of Field Trials—Network Edge Computing Platform," NTT Technical Review, Vol. 20, No. 7, pp. 48–53, July 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202207fa7.html

[4] A. Isomura, N. Shigematsu, I. Ueno, N. Oki, and Y. Arakawa, "Real-time Spatiotemporal Data-management Technology (Axispot™)," NTT Technical Review, Vol. 20, No. 7, pp. 54–60, July 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202207fa8.html

[5] M. Takagi, K. Matsuo, R. Nakada, and K. Mori, "Selective Vehicle-data-collection Algorithm," NTT Technical Review, Vol. 20, No. 7, pp. 61–67, July 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202207fa9.html

[6] K. Matsuo, M. Takagi, R. Nakada, and K. Mori, "Vertically Distributed Computing Technology," NTT Technical Review, Vol. 20, No. 7, pp. 68–75, July 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202207fa10.html

[7] K. Mori, Y. Yokohata, A. Hayashi, T. Hata, and M. Kamiya, "Lane-specific Traffic-jam-detection Technology," NTT Technical Review, Vol. 20, No. 7, pp. 76–82, July 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202207fa11.html

[8] A. Hayashi, Y. Yokohata, T. Hata, K. Mori, and M. Kamiya, "Technology for Calculating Suddenness Index for Aggregated Values," NTT Technical Review, Vol. 20, No. 7, pp. 83–88, July 2022.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202207fa12.html

[9] NTT Topics, "Tech Document on ICT Platform for Connected Vehicles Released by NTT and Toyota," Nov. 29, 2021.
https://group.ntt/en/topics/2021/11/29/techdoc_rel_ntt_toyota.html

Atsushi Koizumi

Senior Manager, R&D Produce Group, NTT Research and Development Planning Department.

He received a B.A. in commerce from Waseda University, Tokyo, in 1995. After working on system development in the public sector at NTT DATA, he was seconded to the NTT DATA Institute of Management Consulting Inc. After returning to NTT DATA, he worked on sales reform and marketing in the Corporate Planning Department then engaged in business planning in the organization dedicated to business process outsourcing and robotic process automation. Currently, he is at NTT and engaged in the research and development of connected vehicles.

Shin Mitsuhashi

Senior Manager, Transport and Mobility Business Producer, Strategic Business Creation Team, NTT Research and Development Planning Department.

He is currently engaged in creating mobility-related services.

Activities and Results of Field Trials—Reference Architecture for a Connected-vehicle Platform

Yu Chiba

Abstract

The NTT Group and Toyota Motor Corporation are collaborating on research and development of an information and communication technology platform for connected vehicles. They conducted joint field trials and verified the platform across a variety of use cases from 2018 to 2020. They also established basic technologies in the course of these trials. This article presents an overview of the reference architecture for the connected-vehicle platform, which collects, stores, and uses controller area network data (vehicle control data) and image data sent from in-vehicle devices. It also reports on the technical results obtained and challenges identified during the implementation of the platform and the field trials.

Keywords: connected vehicles, IoT, big data

1. Characteristics of connected vehicles and technical challenges they pose

As the number of connected vehicles grows rapidly, the amount of data obtained from them, such as controller area network (CAN) data, sensor data, and image data, is growing dramatically. How to process this enormous amount of data efficiently and in real time presents a major technical challenge for a large-scale connected-vehicle platform. Connected vehicles also have several unique characteristics: they need to use a mobile network for communication; move at high speed; have a long life cycle; and the amount of data to be handled varies greatly from hour to hour even within a single day. Therefore, using connected-vehicle data across a variety of use cases poses many complex technical challenges.

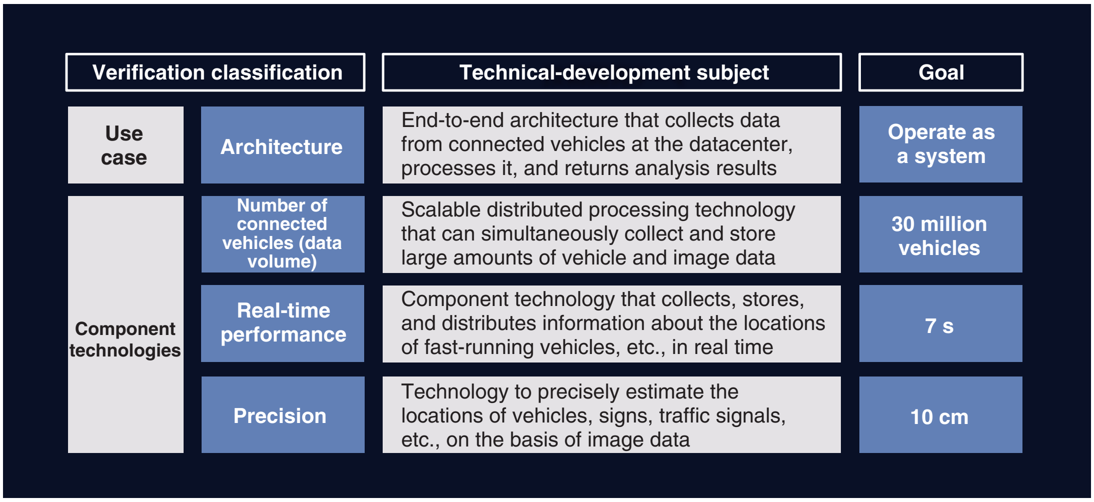
The NTT Group and Toyota Motor Corporation are collaborating on research and development of an information and communication technology platform for connected vehicles. In conducting field trials to verify this platform, we set our goals from the perspectives of three particularly important component technologies: the processing of a large amount of data, real-time performance of the processing of such data, and the degree of precision in the data processing. We have thus studied a connected-vehicle platform that can work across various use cases of connected vehicles and automated driving (**Fig. 1**).

2. Use cases for the field trials

The field trials were conducted using test vehicles on public roads for three years from 2018 to 2020. The main objective of these trials was to establish the technologies for and evaluate the performance of a large-scale platform that will be able to handle millions of connected vehicles in the future. We verified the platform using the following three sample use cases, which are likely to be put into practice (**Fig. 2**).

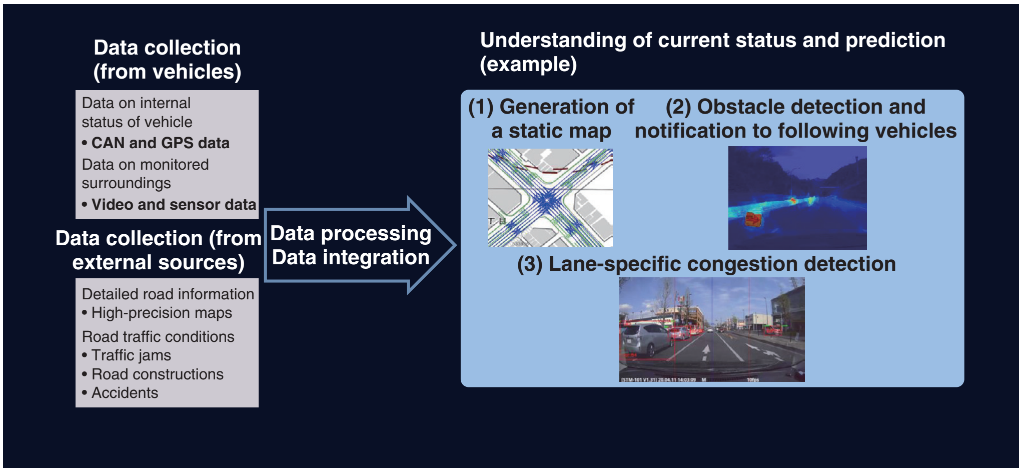
(1) Generation of a static map: The datacenter analyzes vehicle-location data and image data sent from connected vehicles and generates a high-precision static map required for automated driving.

(2) Obstacle detection and notification to following vehicles: Dangerous obstacles, such as falling rocks on the road, are detected using image data from cameras mounted on connected vehicles. The datacenter manages this



Verification classification		Technical-development subject	Goal
Use case	Architecture	End-to-end architecture that collects data from connected vehicles at the datacenter, processes it, and returns analysis results	Operate as a system
Component technologies	Number of connected vehicles (data volume)	Scalable distributed processing technology that can simultaneously collect and store large amounts of vehicle and image data	30 million vehicles
	Real-time performance	Component technology that collects, stores, and distributes information about the locations of fast-running vehicles, etc., in real time	7 s
	Precision	Technology to precisely estimate the locations of vehicles, signs, traffic signals, etc., on the basis of image data	10 cm

Fig. 1. Technical items and target values.



GPS: Global Positioning System

Fig. 2. Sample use cases.

information and notifies the following vehicles of danger.

(3) Lane-specific congestion detection: Both recurring and non-recurring traffic jams on each lane are detected using statistical information in CAN data and real-time image data. The cause of a non-recurring traffic jam is identified by analyzing images of the front point of the traffic jam.

We evaluated the feasibility of these three use cases through the field trials. Among the component technologies needed for a large-scale connected-vehicle platform, we focused on the above three component technologies and set the following goals: the processing of large amounts of data, real-time processing of such data, and the degree of precision in the data processing.

3. Reference architecture for the connected-vehicle platform

Data from connected vehicles are collected via wireless networks (LTE (Long-Term Evolution) network and the 5th-generation mobile communication

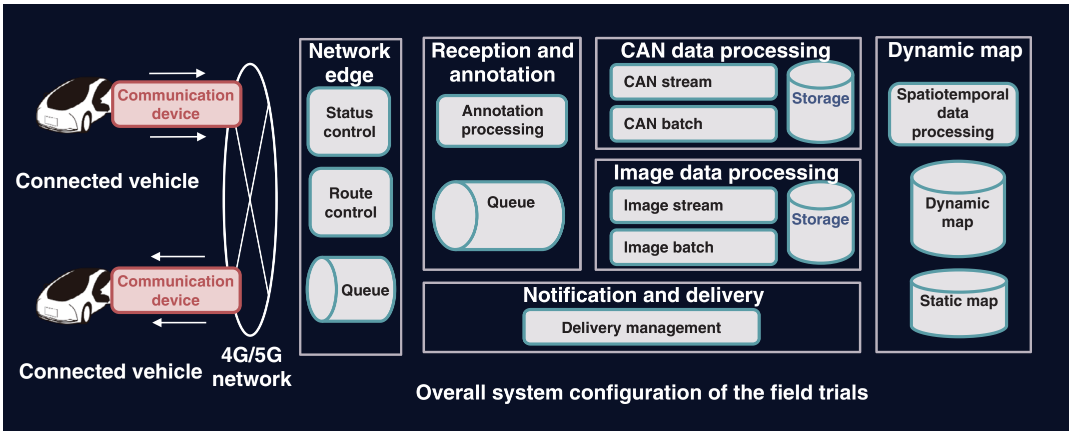


Fig. 3. Reference architecture.

system (5G) network) and stored in servers in datacenters. The data are analyzed, and analysis results are sent back to connected vehicles. We incorporated this end-to-end system into the reference architecture for the connected-vehicle platform. This reference architecture consists of six platforms: a *network-edge computing platform*, which has a real-time link with connected vehicles, a *reception and annotation platform*, which receives data from edge computing nodes, *CAN data processing platform* and *image data processing platform*, both of which process data, *dynamic map platform*, which manages data, and *notification and delivery platform*, which manages communication from datacenters to connected vehicles.

We implemented this architecture using open-source software programs that had become de facto standards in the global market. The goal of this collaboration is to develop technologies that will lead to standardization with a policy of establishing technologies that are more open and not dependent on the proprietary software of a specific company. Therefore, the software is structured in such a way that it can be revised as technology advances and innovations emerge (**Fig. 3**).

4. Verification of use cases

A server environment for the field trials was implemented in a datacenter based on the reference architecture described above. The feasibility of the following three use cases was verified using test vehicles.

(1) Generation of a static map

In this use case, an operator at the datacenter first sends an instruction to connected vehicles to generate a map. When a connected vehicle running in the target area receives the instruction, it sends CAN data and image data to the datacenter. The datacenter executes preprocessing for generating a map from the image data, estimates the locations of traffic lights and other objects from the image data, generates map data, and registers the data in the map database. We verified all these processes (**Fig. 4**).

(2) Obstacle detection and notification to following vehicles

In this use case, the datacenter uses image data from onboard cameras and learned obstacle data to infer an object in images. If the object is an obstacle, its location is estimated and registered in the dynamic map database. A challenge in this use case is real-time performance. For this to be practical, it is necessary to detect an obstacle and notify the vehicles approaching in the rear of it within 7 seconds. We adopted an architecture that offloads parts of this processing to the network-edge computing nodes, which greatly improved real-time performance. We further improved this performance by sending danger notifications in two stages. The first notification is sent promptly to the vehicles in a wider area without taking time to narrow down the affected area. The vehicle in danger in a specific lane is then identified and a second notification is sent to that vehicle (**Fig. 5**).

(3) Lane-specific congestion detection

In this use case, the datacenter identifies the cause of a traffic jam in three steps. In Step 1, the datacenter analyzes CAN data of running vehicles in real time and narrows down the candidate traffic-jam front

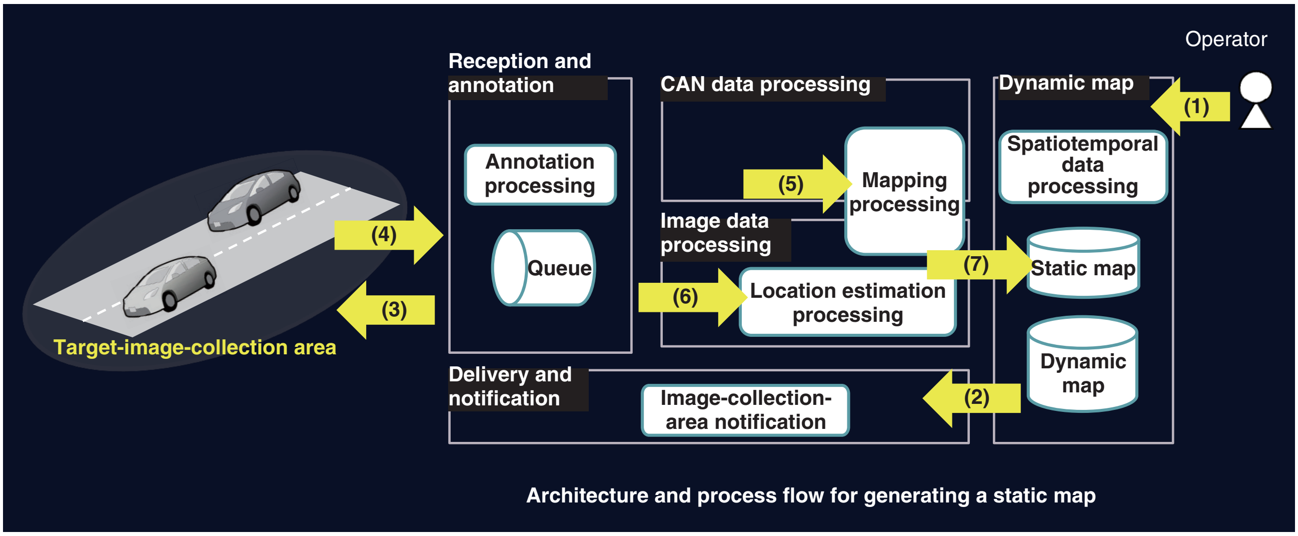


Fig. 4. Generation of a static map.

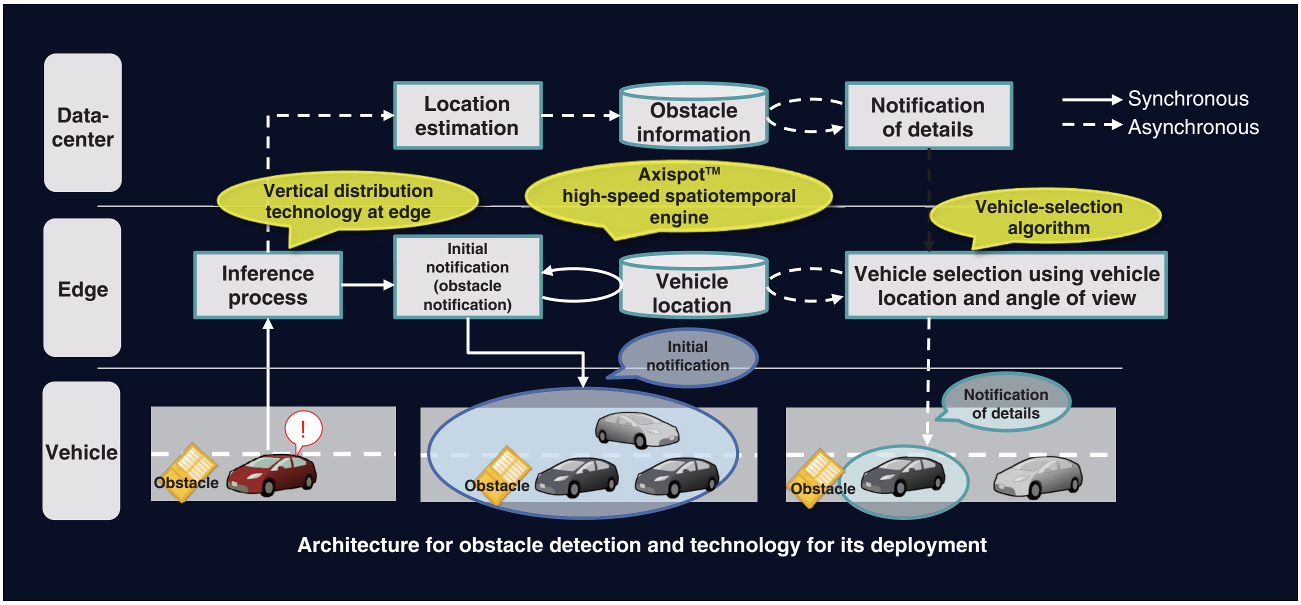


Fig. 5. Obstacle detection and notification to following vehicles.

point on the basis of vehicle density per mesh and per lane. Recurring traffic jams are excluded because their causes can be identified in advance; thus, we focused on non-recurring traffic jams. In Step 2, the datacenter collects image data from the vehicles running in the vicinity of the potentially congested lane identified in Step 1. Using the image data, the datacenter determines whether there is a traffic jam, its location, and its front point, and identifies its cause. In Step 3, the datacenter notifies the vehicles running in the affected lane of the cause of the traffic jam obtained in Step 2 (**Fig. 6**).

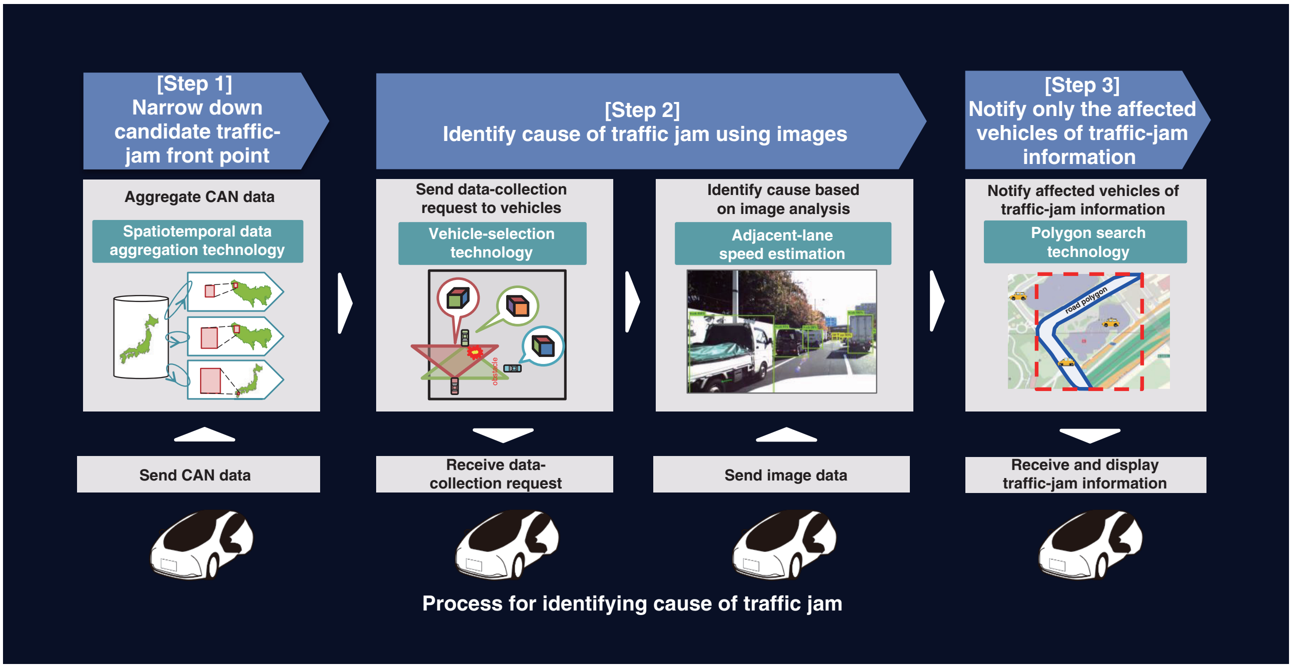


Fig. 6. Lane-specific congestion detection.

5. Verification of the platforms of the reference architecture

We evaluated the following platforms of the reference architecture.

(1) Reception and annotation

We evaluated this platform by focusing on the performance limits of CAN data and image data processing. Specifically, we evaluated the load performance of annotation processing and queuing processing and identified bottlenecks under a simulated condition in which 5 million vehicles were sending data at intervals ranging from a minimum of 1 second up to 10 seconds.

(2) CAN data processing

We evaluated this platform in terms of real-time performance and throughput. We implemented real-time processes, such as the regular storage of vehicle location information, through CAN stream processing and evaluated the performance limit of their respective response times.

(3) Image data processing

We evaluated this platform in terms of real-time performance and throughput. We implemented real-time image processes, such as obstacle detection, through image stream processing and evaluated the performance limit of their respective response times.

(4) Notification and delivery processing

We evaluated this platform in terms of the cost of image collection, which can become a bottleneck in each use case. An algorithm for selective collection of vehicle data, which was developed by NTT, was used to reduce the amount of image data collected. We evaluated the performance of communication infrastructure technologies assuming that this algorithm was used.

The verification of the network-edge computing platform is not described here because it is described in another article in this issue: "Activities and Results of Field Trials—Network Edge Computing Platform" [1].

6. Issues identified and future activities

We verified the reference architecture for three sample use cases using test vehicles running on public roads. In evaluating the communication infrastructure technologies needed for a large-scale connected-vehicle platform, we set goals for data volume, real-time performance, and precision (**Fig. 7**). Regarding data volume, we used simulation data and successfully verified that the data processing platform can handle 30 million vehicles. Regarding real-time performance, we used network-edge computing and

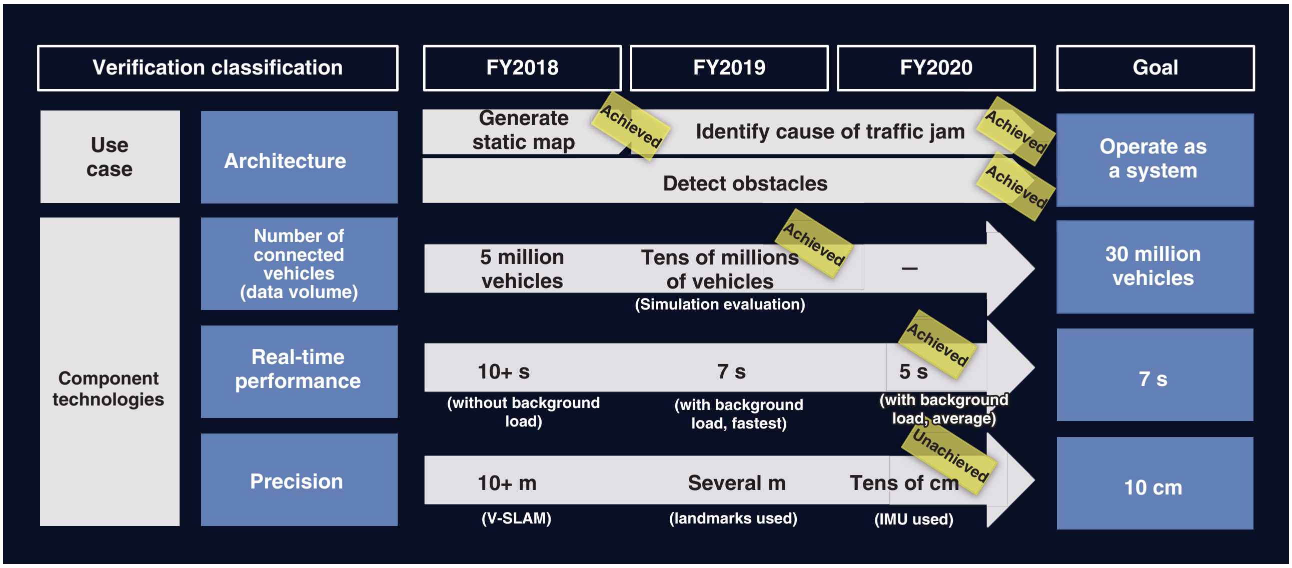


Fig. 7. Activity results.

achieved an average notification time (time taken before sending notification) of 5 seconds (the target was within 7 seconds) in the obstacle-detection use case. We attempted to improve precision by using visual simultaneous localization and mapping (V-SLAM)*, landmark location information, and an inertial measurement unit (IMU). Despite these efforts, we were not able to achieve our target of 10-centimeter precision. This will be an issue for future technological development.

While we achieved our initial goals, except for precision, we also identified a number of technical issues. The collection, storage, and utilization of large amounts of image and other data impose a heavy burden on networks and server resources. For these technologies to be implemented in society, it is necessary to develop technologies that are not only functionally feasible but also more efficient and less expensive. To achieve these targets, we are developing technologies that use edge computing to distribute processing loads and use resources efficiently. We will also expand the application of these technologies to cases in which they can address social issues, such as contributing to a low-carbon society, a contribution that is particularly expected of the automobile industry.

Reference

[1] R. Noji, T. Kamei, K. Funabiki, D. Murata, and Y. Kanemaru, “Activities and Results of Field Trials—Network Edge Computing Platform,” NTT Technical Review, Vol. 20, No. 7, pp. 48–53, July 2022. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202207fa7.html

Yu Chiba
Senior Manager, Manufacturing IT Innovation Division, NTT DATA Corporation.

He received an M.S. in computer science from Tohoku University, Miyagi, in 2005 and joined NTT DATA the same year. He was involved in the research and development of software for large-scale systems in the financial sector as well as the development of many systems in public and enterprise sectors. After working as a branding manager in the Public Relations Department, he has been in charge of technological development of a connected-vehicle platform since 2018.

---

* V-SLAM: A technology that simultaneously determines the system’s location and generates an environmental map based on three-dimensional information of its surroundings.

Activities and Results of Field Trials—Network Edge Computing Platform

Ryosuke Noji, Takayuki Kamei, Kaito Funabiki, Daisuke Murata, and Yuken Kanemaru

Abstract

Toyota Motor Corporation and the NTT Group conducted field trials to verify an information and communication technology platform for connected vehicles over a three-year period beginning in 2018 and shared their respective technologies and expertise regarding connected vehicles. Distributed processing using edge nodes increased the efficiency and processing speed of the platform. However, the distribution of processing sites created new challenges. To address these challenges, we developed an architecture in which multiple network functions are allocated to edge nodes and verified its effectiveness through these field trials.

Keywords: network, edge node, distributed processing

1. Introduction

At the early stage of the field trials conducted by Toyota Motor Corporation and the NTT Group, a large datacenter collected data from vehicles, and many servers were installed at the datacenter to process the data in a distributed manner. However, this architecture was not able to achieve the target scalability (handling 30 million vehicles) and processing time (sending a response within 7 seconds). To overcome this limitation, edge nodes were placed between vehicles and the datacenter, and some data processing was delegated to these edge nodes. Usually, the term "edge node" refers to a terminal device, for example, the computing resource in a vehicle in the case of a connected vehicle. With our new approach, however, edge nodes are also geographically distributed but located between vehicles and the datacenter.

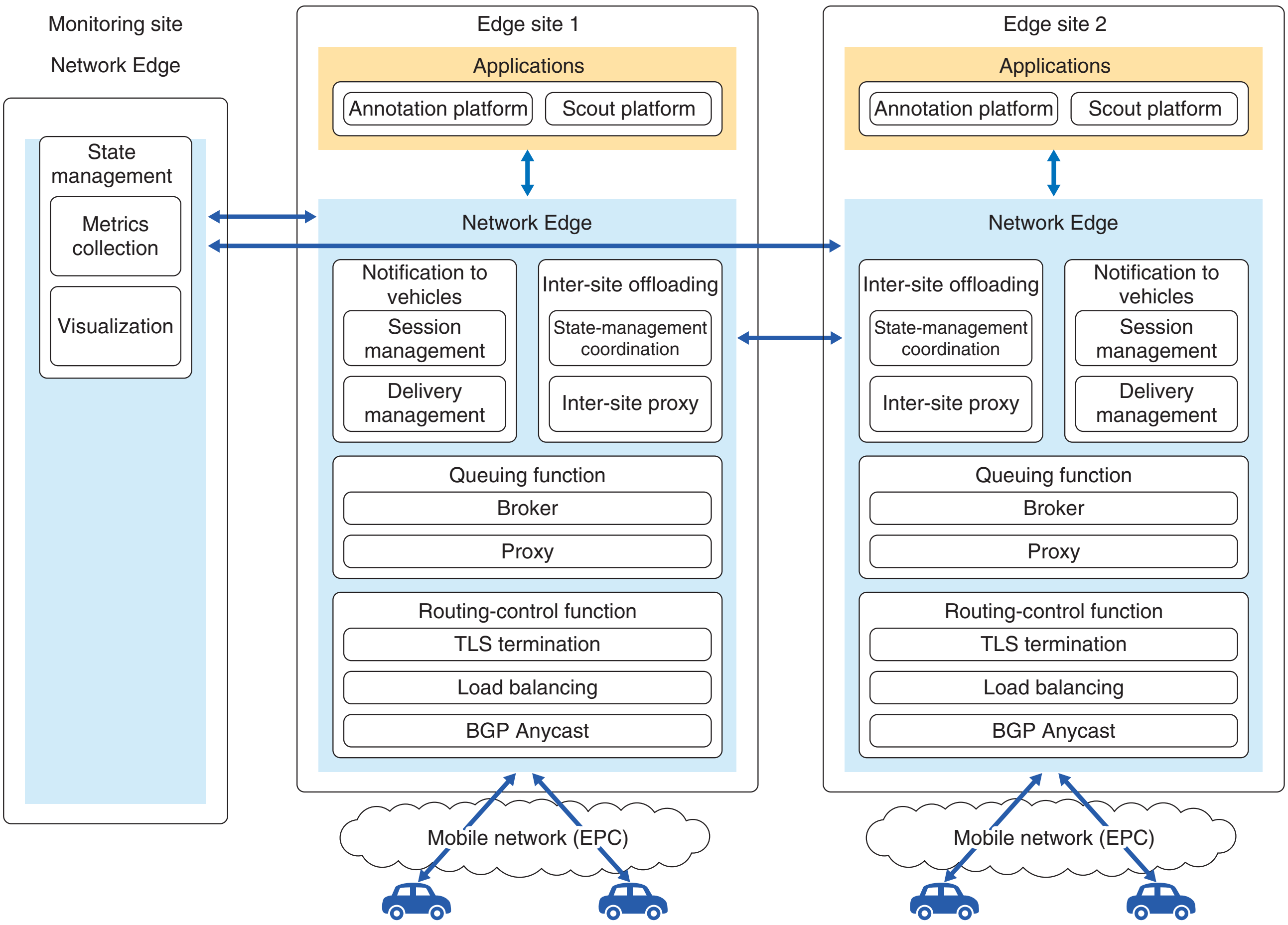
Challenges in using edge nodes include not only the need to establish the appropriate application architecture but also a network-related need to transport data appropriately so that load balancing can be achieved between the servers and distributed edge nodes. To address the latter issue, we developed an architecture in which processing systems executing multiple network functions are located between a vehicle and the application group. In the field trials, we verified whether this architecture is effective for solving this issue. To distinguish these processing systems from applications installed at terminals (edges), we refer to these processing systems as Network Edge (**Fig. 1**).

Network Edge acts as a gateway when vehicles upload data and a gateway when applications in the datacenter send notifications to vehicles. One of the goals with Network Edge is to hide the complexity of the network or infrastructure so that application developers can concentrate on developing application functions.

2. Target use cases

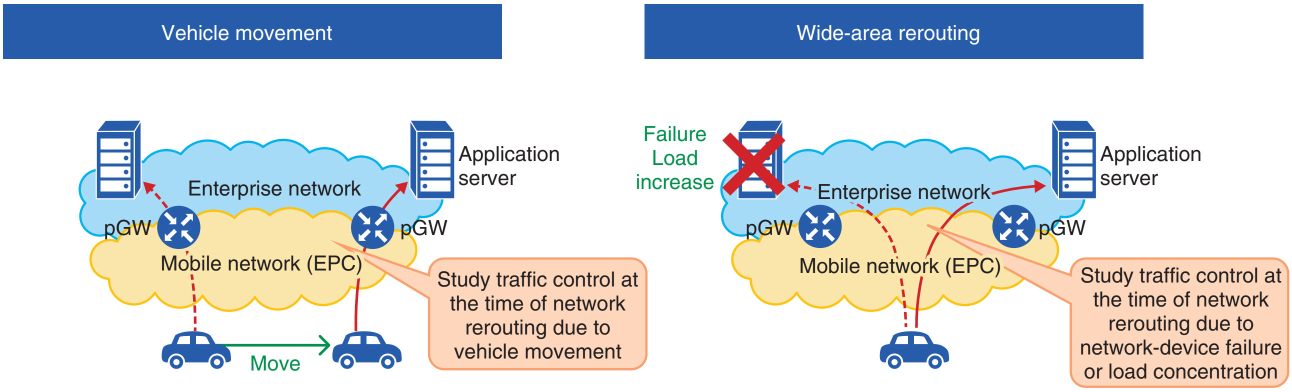
To avoid narrowing down target use cases too much, we selected two fundamental use cases to verify the performance of Network Edge: vehicle movement and wide-area rerouting (**Fig. 2**).

In the use case of vehicle movement, the application



BGP: Border Gateway Protocol
EPC: Evolved Packet Core
TLS: Transport Layer Security

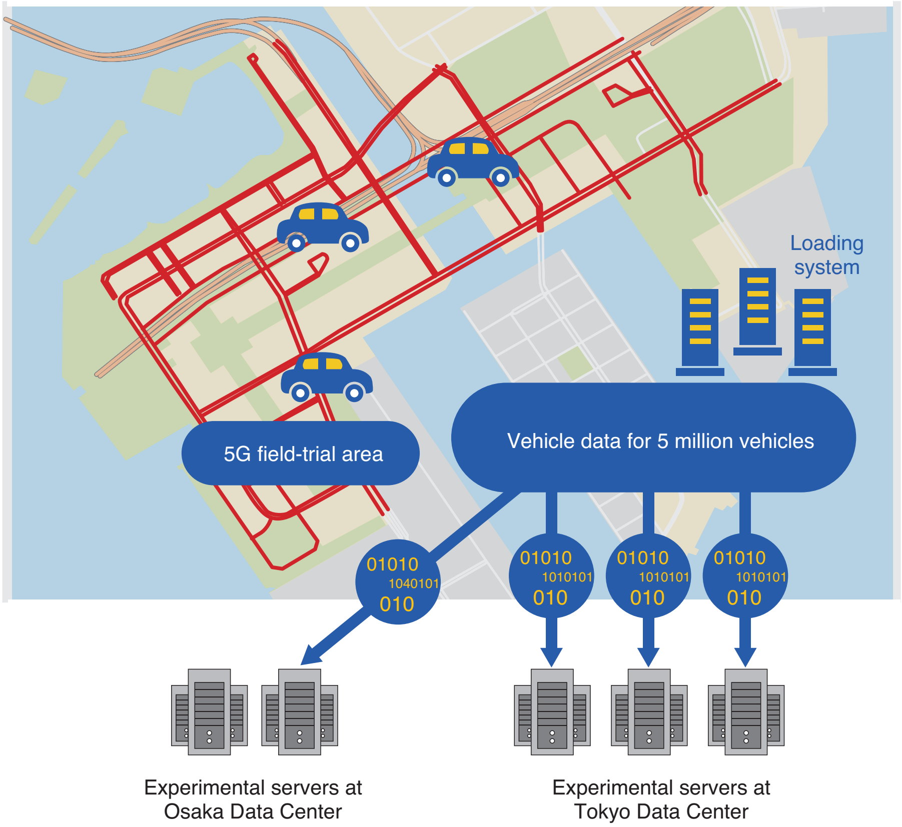
Fig. 1. Overview of Network Edge.



pGW: Packet Data Network Gateway

Fig. 2. Fundamental use cases.

The field trials using actual vehicles were conducted in the Odaiba area in Tokyo.



5G: the 5th-generation mobile communication system

Fig. 3. Overview of the environment for the field trials.

facility and server to which a vehicle is connected are switched as the vehicle moves. In principle, data are processed at the edge node closest to the vehicle concerned, and the results are sent from there to neighboring vehicles. This can be described as *local production for local consumption of data*. As the vehicle moves, it is necessary to switch not only the edge node to which the vehicle is connected but also the application executing the server. We verified whether this switching requirement can be supported by network technology.

In the use case of wide-area rerouting, the application facility to which a vehicle is connected is switched when a failure occurs at the facility or when it is necessary to execute load balancing. As edge nodes are distributed, a failure or overload of an edge node makes it necessary to transfer the processing load from the edge node concerned to neighboring edge nodes. We verified whether, in the proposed architecture, failures of edge nodes can be detected and transfer of processing loads between edge nodes can be supported.

3. Network Edge functions and verification items

Network Edge performs three main functions: control of routing to application servers at datacenters, load balancing based on metrics, and message queuing distributed over a wide area. If we put geographic distribution aside for the moment, data from a vehicle reaches an edge node via a wireless access network, such as a mobile network. The control functions mentioned above send data to applications on the edge node or, in some cases, to applications in the datacenter site.

In the field trials, we implemented an environment with multiple edge nodes using datacenters in Tokyo and Osaka to confirm the two use cases and verified the effectiveness of the three functions (**Fig. 3**). Since the use of Transport Layer Security (TLS) is assumed to encrypt communication with vehicles for increased security, terminating TLS at Network Edge may become a processing bottleneck. Therefore, we verified the performance of TLS termination. Details of these verifications are described below.

4. Routing control

We investigated two methods of controlling routing data from a vehicle to an application server: the domain name system (DNS) method and load balancer (LB) method. We confirmed the basic functions and performance of the two methods and investigated how they behave when a vehicle moves or wide-area rerouting takes place as a result of a facility failure.

(1) DNS method: Border Gateway Protocol (BGP) Anycast connects a vehicle to the nearest DNS server, and the DNS server selects the most appropriate datacenter and application server and notifies the vehicle of this information.

(2) LB method: BGP Anycast connects a vehicle to the nearest LB server, and the LB forwards requests from vehicles to the most appropriate server at the most appropriate datacenter.

With the LB method, vehicles always communicate with the application via the LB. This lowers performance but makes finer-granularity control possible. We compared the two methods in the early stage of the field trials and found that both were applicable to the target use cases. Therefore, from then on, the field trials were based on the LB method.

5. Load balancing using metrics

On the assumption that applications were located in multiple servers in multiple datacenters and processed in a distributed manner, we verified load balancing in which the datacenter and server to use were determined on the basis of metrics information, and in which the routing control described above was used.

We examined whether extreme performance degradation could be avoided by detecting server or datacenter failures, software process failures, or overloads from received information about response times from application servers, facility information, and metrics of the infrastructure resources, and by conducting load balancing on the basis of the detected situation. We also verified whether the load can be distributed over a wide area in specified proportions. For example, when the application load exceeds a threshold during an experiment, 70% of the load at Center 1 is processed at Center 1 and the remaining 30% is offloaded to Center 2.

6. Wide-area distributed message queuing

If the only response to a server or datacenter failure is to reroute data to a server in another datacenter, data transmission from vehicles would fail during the rerouting time. If this occurs, depending on the situation, it is necessary either to retransmit data or give up sending data. The purpose of the wide-area distributed message-queuing function is to prevent this from occurring.

As the term "message queue" indicates, the message-queuing function receives data only temporarily. A message queue resides at each edge node; therefore, message queues at different nodes work together to deliver data to the appropriate servers when offloading processing loads.

Since the edge node to which a vehicle is connected is switched as the vehicle moves, the datacenter that wants to send a notification to a vehicle needs to know the edge node to which the vehicle is currently connected. We aimed to satisfy this need using this queuing function. When the datacenter receives messages from applications, it retains them so that it can send them to the vehicle concerned even if the vehicle has moved to an area covered by a different edge node. In the field trials, we examined whether this occurs reliably.

If Apache Kafka or similar software is used to execute these processes, problems arising in that data are exchanged unnecessarily between datacenters; thus, it is impossible to manage notification delivery. To avoid this problem, we combined basic message-queuing software with a proxy program we developed. (We used NATS, which is open-source software, for the basic message-queuing function.) The properties required for data upload and data download (sending notifications) are different. Therefore, for the former, we used a mechanism with which the target data are first selected then sent. For the latter, we adopted a mechanism with which metadata are shared and there is a logical queue that spans different datacenters.

7. Verification of TLS performance

Vehicles communicate using TLS to ensure security. Applications communicate using HTTPS (Hypertext

Transfer Protocol Secure) or MQTTS (Message Queuing Telemetry Transport Secure) because they use HTTP or MQTT. Since we chose to use the LB method for routing control in this verification, the Layer 7 LB software needs to handle a large number of transactions. Normal tuning alone would result in a central processing unit bottleneck. Therefore, TLS processing was offloaded to another piece of hardware. We verified whether this would improve performance and resource-usage efficiency. Two types of hardware, TLS Accelerator and SmartNIC, both of which can be installed in a standard AMD64 server, were used for this verification. Although we have not yet been able to confirm the effectiveness of this method when the dominant traffic is data uploaded from vehicles, we will continue to verify this method since, for a connected vehicle platform designed for tens of millions of vehicles, it is important to minimize the number of required servers by using resources efficiently.

8. Future outlook

In these field trials, we implemented an architecture in which Network Edge, which perform multiple functions, is placed between vehicles and applications [1] and were able to confirm the effectiveness of this architecture.

On the basis of this architecture, we will further develop technologies and attempt to combine them with other technologies to develop a more advanced platform. For example, we are considering the application of artificial intelligence to metrics-based load balancing and coordination with priority control of application servers. We will continue our efforts to achieve a world in which intelligent network technology supports real-time communication of tens of millions of fast-moving vehicles and in which engineers can concentrate on developing applications without worrying about complex infrastructure conditions.

Reference

[1] Press release issued by NTT, “Tech Document on ICT Platform for Connected Vehicles Released by NTT and Toyota,” Nov. 29, 2021. https://group.ntt/en/topics/2021/11/29/techdoc_rel_ntt_toyota.html

Ryosuke Noji

Senior Manager, Smart Mobility, Smart World Business, Business Solution Division, NTT Communications Corporation.

He started his career as a solution engineer at NTT Communications in 2007. After several projects, he joined the research and development department. As a data platform architect, he designed and developed data platforms for the company's IPTV Service and the Data Science Lab, which is a data science Center of Excellence at NTT Communications, and analyzed and developed algorithms for several services. He has been engaged in solution development for the automotive industry since 2018 and has led a connected-vehicle project for developing network and computing services or solutions, including edge computing, for connected-vehicle big data.

Takayuki Kamei

Engineer, Dept. Technology, Innovation Center, NTT Communications Corporation.

He started his career as a network engineer at NTT Communications in 2013. He has continued to research and develop network functions virtualization, software-defined wide area network, application-layer overlay networks, and content delivery network. He has been engaged in the development of edge computing and traffic optimization since 2019 and led the architecture design of distributed platforms over a WAN including connected-vehicle use cases.

Kaito Funabiki

Engineer, Smart Mobility, Smart World Business, Business Solution Division, NTT Communications Corporation.

He started his career as a network engineer at NTT Communications in 2017. After working on the operation and development of an internet exchange service, he has been engaged in a connected-vehicle project for developing network and computing services or solutions, including edge computing, for connected-vehicle big data since 2020.

Daisuke Murata

Engineer, Innovation Center, NTT Communications Corporation.

He received an M.E. in information science from Nara Institute of Science and Technology in 2014. In 2018, he joined NTT Communications and has been engaged in the development/research of computer systems and computer networking technologies, especially in the area of TLS acceleration since 2020.

Yuken Kanemaru

Engineer, Dept. Technology, Innovation Center, NTT Communications Corporation.

He started his career as a network engineer at NTT Communications in 2019. He has continued to research and develop application-layer overlay networks and content delivery network. He has been engaged in the development of edge computing, traffic optimization, and distributed platforms over a WAN including connected-vehicle use cases.

Real-time Spatiotemporal Data-management Technology (Axispot™)

Atsushi Isomura, Naoko Shigematsu, Isoo Ueno, Nobuhiro Oki, and Yutaka Arakawa

Abstract

To enable the *obstacle-detection use case*, in which vehicles approaching an obstacle (e.g., falling object) on the road are immediately notified of the obstacle, and the *lane-specific congestion detection use case*, in which the number of vehicles in each lane on the road is determined, it is necessary to store data sent simultaneously from a large number of connected vehicles, search the data in real time for vehicles present within a certain area (mesh area, road, parking lot, etc.) at any specific time, and determine the number of these vehicles. This article describes the real-time spatiotemporal data-management technology (Axispot™) that we are developing to meet these requirements.

Keywords: digital twin, spatiotemporal data, Axispot™

1. Introduction

Internet of Things (IoT) technology makes it possible to collect a range of information about people, objects, and the natural environment in real space and centrally manage these pieces of information on the cloud. IoT is becoming essential for next-generation services that require management of moving objects, such as people and vehicles. Such services include lane-specific congestion detection, route search to avoid congested routes, and aerial delivery of goods by drones. In a collaboration between NTT and Toyota Motor Corporation, we are studying an obstacle-detection use case, in which vehicles approaching an obstacle on the road are immediately notified of the obstacle. To enable this use case, it is necessary to store a large amount of data in real time, search the data in real time for vehicles present within a certain area at any specific time, and determine the number of these vehicles.

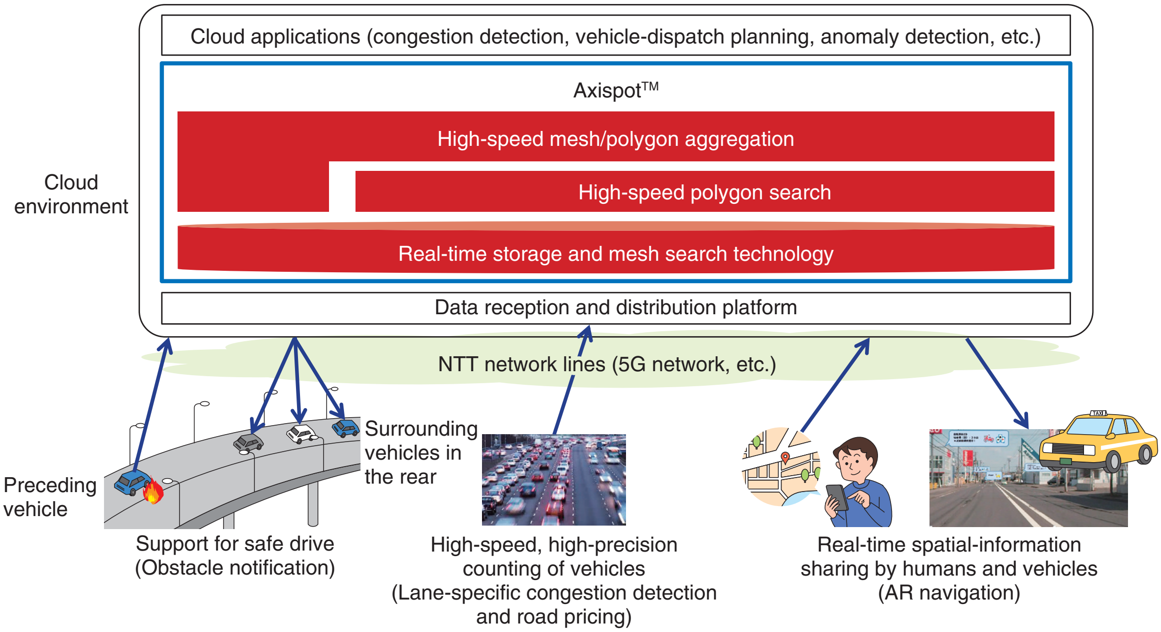
NTT Human Informatics Laboratories is developing a real-time spatiotemporal data-management system, Axispot™[*1] [1, 2], which stores data sent simultaneously from a large number of moving objects and searches the data in real time for moving objects present within a certain area at any specific time (**Fig. 1**). The following sections present four technical issues facing current technologies and introduce the four core technologies that are included in Axispot™, which are intended to overcome these technical issues. We also evaluate the improvement in performance achieved with these technologies. Finally, the issues being addressed to further advance our research and development efforts, as well as their future outlook, are described.

2. Technical issues

A spatiotemporal database is a database for efficiently storing, searching for, and extracting groups of data related to both spatial information, such as latitude and longitude, and temporal information, such as time and period [3, 4]. Searching a spatiotemporal

---

[*1] Axispot™: A spatiotemporal data-management technology that identifies (spots) data from multi-dimensional axes in real time or a system using the technology.



AR: augmented reality
5G: fifth-generation mobile communications

Fig. 1. Image of using Axispot™.

database for objects present within a certain area of a certain shape (road, parking lot, etc.) using spatial information is called a polygon search. To enable the obstacle-detection use case, it is necessary to store data generated by a huge number of vehicles throughout the country in a spatiotemporal database and use polygon searches to extract only specific vehicles. However, the following four technical issues stand in the way of achieving our target performance.

- Issue 1: In an ordinary spatiotemporal database, indexes are generated on the basis of a search tree[*2], which is commonly used in relational databases. However, when data are to be stored or deleted in such a database, an update to the search tree occurs. If the database is handling data from tens of millions of vehicles, such search tree updates significantly reduce the efficiency of storing, updating, or deleting data, making it difficult to store data in real time.
- Issue 2: When it is necessary to handle a large amount of data, a distributed database consisting of multiple servers is used. In such a case, each server is usually assigned to the management of data for a specific region. However, the number of vehicles on the road differs from region to region and varies with the time of day. Therefore, the amount of data handled by a server varies greatly among the servers that make up a distributed database. As a result, the load may be concentrated on a few servers, thereby significantly reducing processing performance and making real-time storage and search difficult.
- Issue 3: With a polygon search, the more vertices the target polygon has, the slower the search becomes [5]. Unfortunately, the polygons provided by high-precision maps[*3], which have become popular, have a large number of vertices. Therefore, polygon searches that use polygons provided by a high-precision map are slow, falling far short of the speed required for real-time search.
- Issue 4: When aggregating the number of vehicles on each road around the country on the basis of spatiotemporal data in a database, it is necessary to assign the number of meshes or polygons within the aggregation area to computers in a manner that takes into account the number of running vehicles in each mesh or polygon. In

---

[*2] Search tree: Indexing technology for data indexes that have a tree structure such as an r-tree.

[*3] High-precision map: High-precision map data that include information on each lane of each road.

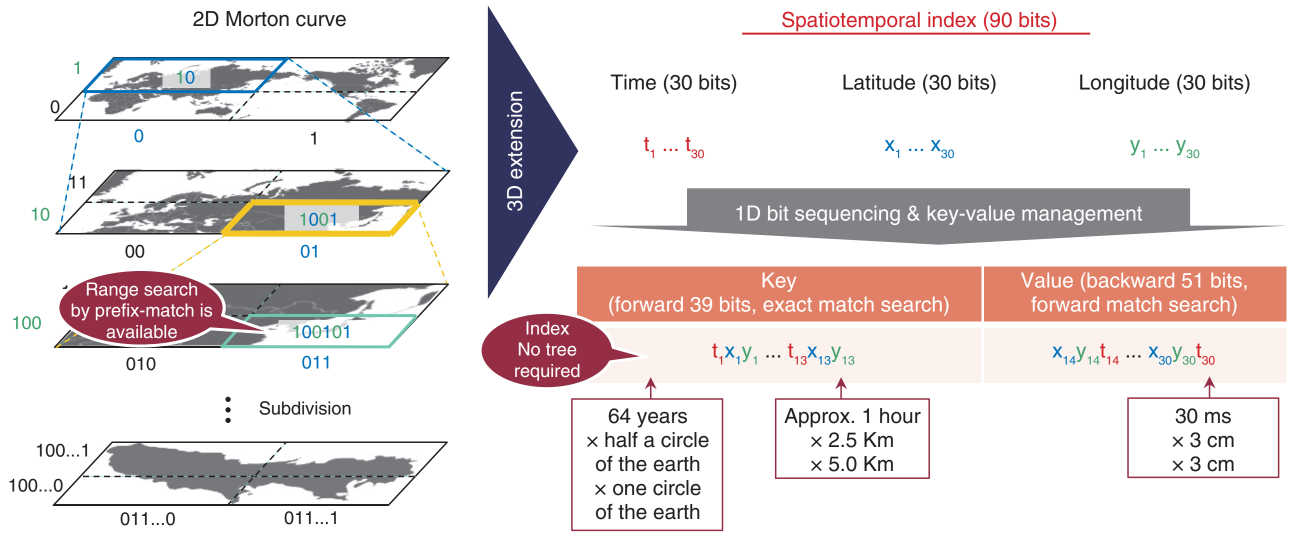


Fig. 2. Spatiotemporal index.

most services, the number of meshes or polygons is manually assigned to computers in advance. However, as mentioned earlier, the number of running vehicles varies depending on the region and time of day. Consequently, the amount of jobs[*4] allocated to each computer is highly uneven, making aggregation computation inefficient.

3. Proposed technologies

This section describes the four technologies proposed by NTT Human Informatics Laboratories to address technical issues (1) through (4) mentioned above then introduces the overall architecture of Axispot™.

3.1 Spatiotemporal indexing

This technology does not use a tree structure. It is an indexing technology that uses the characteristics of a space-filling curve called a Morton curve. The two-dimensional (2D) Morton curve shown on the left of **Fig. 2** is a process of repeatedly dividing the world map into four parts and assigning either bit [0] or [1] to each partitioned area. The longer this 1D bit sequence becomes, the narrower the area it represents. A forward match search can be executed on this bit sequence to search a certain range of space. For example, a search of the forward bit [1001...] is a search of the yellow box area in the world map shown in Fig. 2. We added a time dimension to the Morton curve, thereby extending it to a 3D Morton curve. Furthermore, the forward and backward bits of the generated bit sequence are respectively stored as Key and Value in a database called key-value store, as shown on the right of Fig. 2. As a result, a spatiotemporal range search can be executed on the basis of an exact match of Key. This spatiotemporal indexing technology makes both high-speed data storage and real-time spatiotemporal range search possible without using a tree structure.

3.2 Limited node selection

This technology is a new data-distribution technology that determines multiple combinations of servers on the basis of location and time and distributes data to one of the combinations. This technology provides the three effects shown in **Fig. 3**. (1) Load distribution can be equalized by selecting multiple servers instead of selecting one specific server for a specific region. (2) Since the selected servers included in a combination change with the passage of time, continuous load on specific servers can be prevented. (3) At the time of the search, a unique combination of servers can be identified on the basis of time and space. Therefore, unnecessary searches involving all servers can be prevented. This limited node-selection technology prevents load concentration on specific servers and enables high-speed storage and search.

[*4] Job: A group of processing operations.

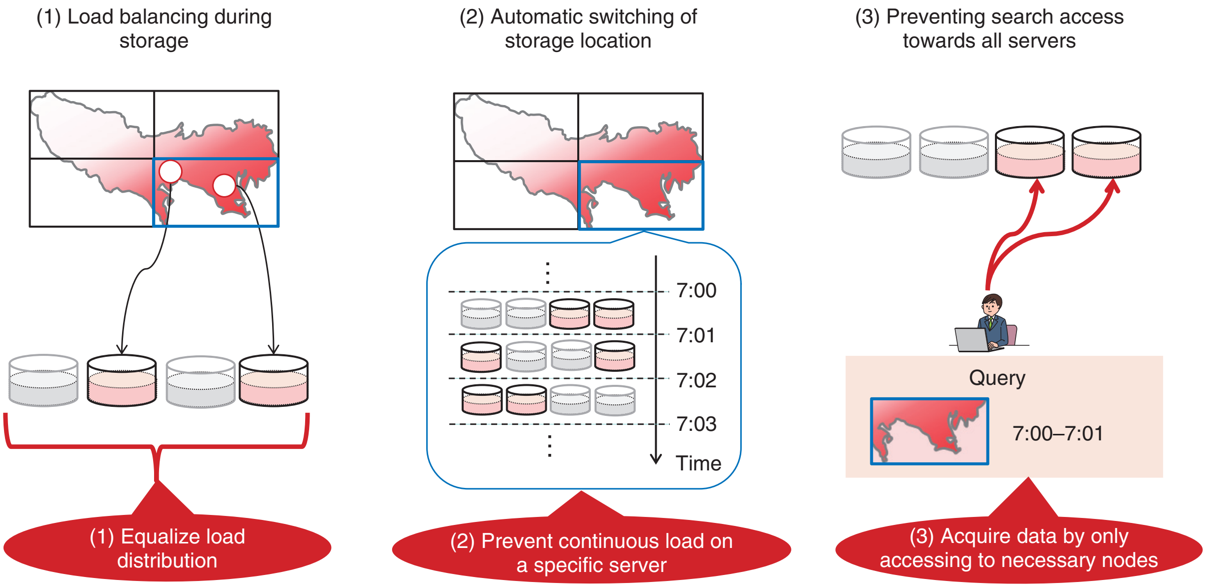


Fig. 3. Limited node selection.

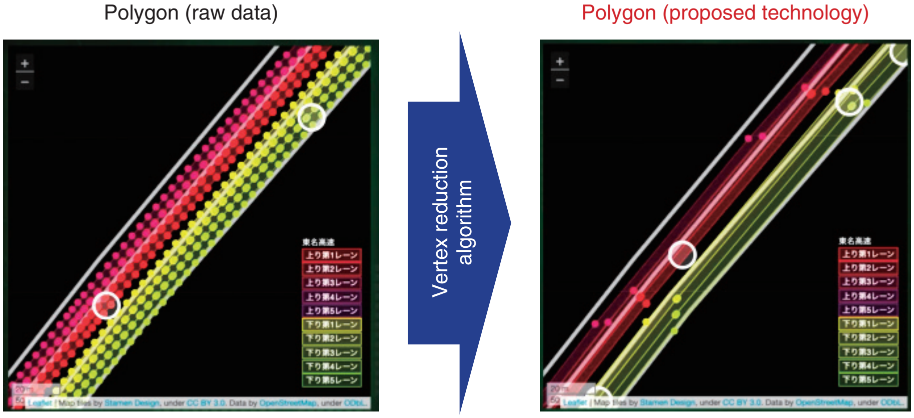


Fig. 4. Fast polygon search with vertex reduction.

3.3 High-speed polygon search

This technology applies the Douglas-Peucker algorithm [6] for simplifying polygonal lines forming polygonal shapes to significantly reduce the number of vertices while maintaining the original polygonal shape and to speed up polygon search. By applying the Douglas-Peucker algorithm to the polygon (raw data) in **Fig. 4**, a polygon with a significantly reduced number of vertices (proposed technology) is generated. After applying this conversion, the high-speed polygon search technology efficiently manages polygons in a database. This technology prevents any reduction in polygon search speed by using polygons generated on the basis of a high-precision map, making real-time search possible.

3.4 High-speed spatiotemporal data aggregation

This technology automatically allocates aggregation jobs within the target aggregation range of spatiotemporal data to computers by using a distributed processing platform. As shown in **Fig. 5**, when data for a spatial area are to be aggregated using meshes, jobs are allocated to computers in such a way that the amount of jobs allocated to each computer is equal

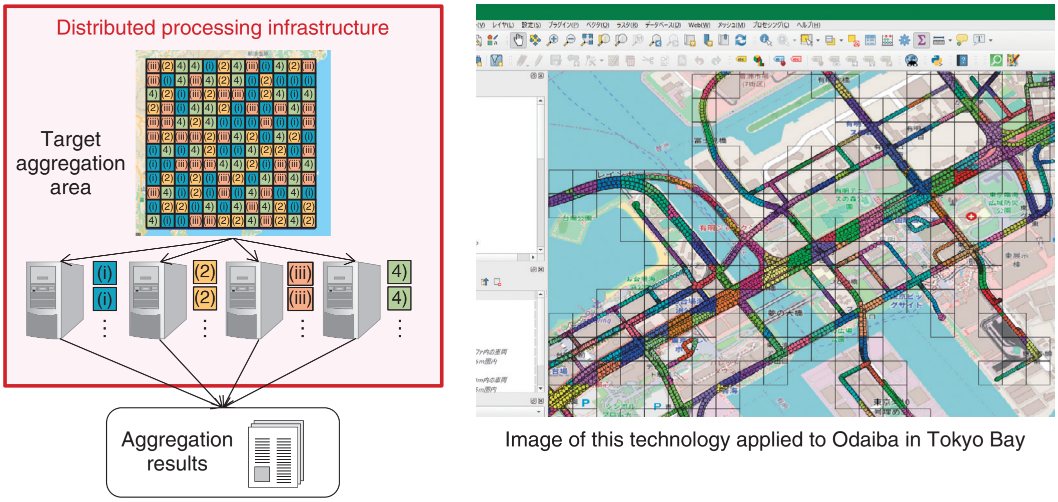


Image of this technology applied to Odaiba in Tokyo Bay

Fig. 5. High-speed mesh-based aggregation.

even though the meshes allocated to each computer may be spatially scattered. This high-speed spatiotemporal data aggregation technology prevents the allocation of uneven amounts of jobs across computers and reduces the aggregation response time.

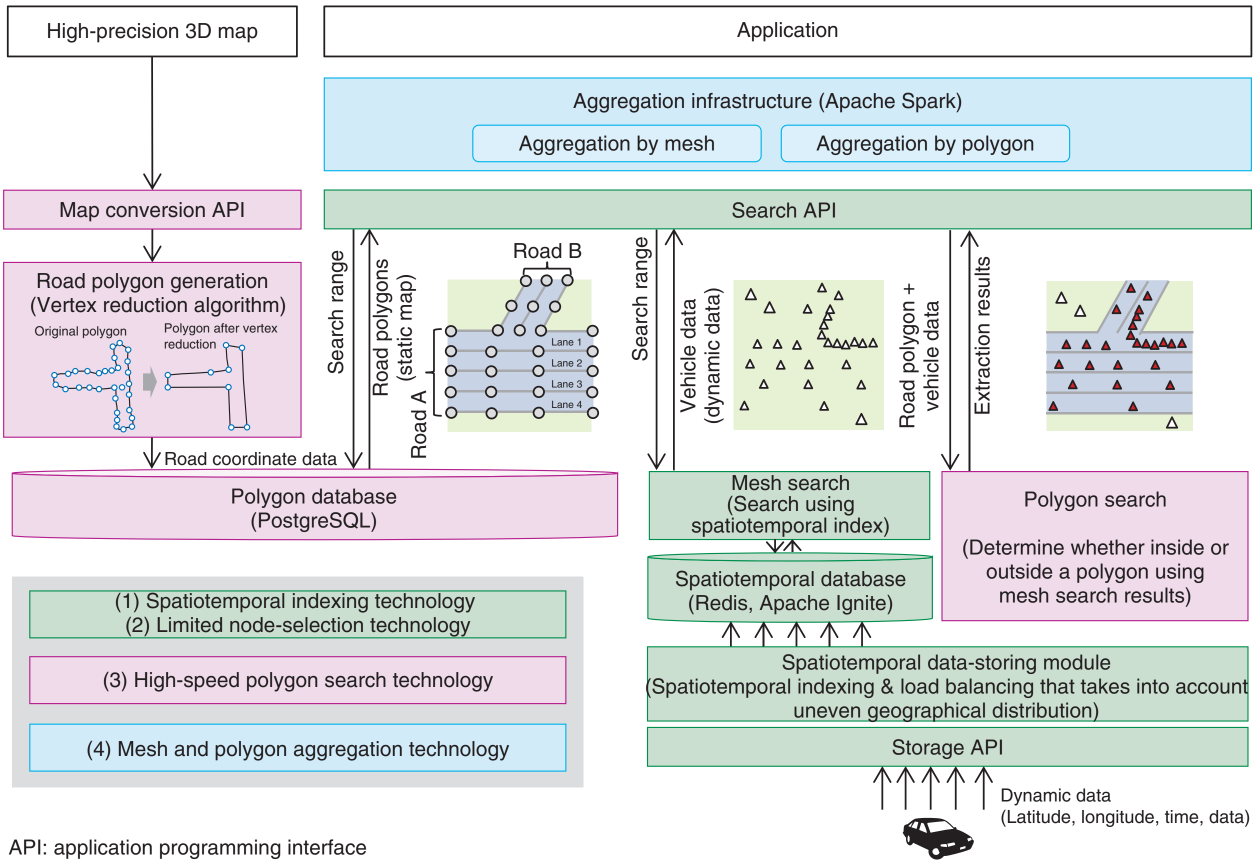
Axispot™ includes these four technologies and has the architecture shown in **Fig. 6**. The spatiotemporal database shown in the figure can use Redis [7] and Apache Ignite [8] to store spatiotemporal data. The static polygon database can use open-source software, such as PostgreSQL [9], to store polygon information.

4. Evaluation

This section shows a comparison between when both spatiotemporal indexing technology and limited node-selection technology are applied to a database and when neither technology is applied (i.e., only conventional technologies are used) to the same database. We collected data for several million vehicles, the number of vehicles constantly running in the Kanto region surrounding Tokyo, and used a dataset that was designed to search for vehicles that exist in a specific space and time. As shown in **Fig. 7(a)**, we confirmed that storage performance improved by a factor of 3.1 and search performance by a factor of 246. As shown in **Fig. 7(b)**, we also confirmed that these technologies prevent load concentration because they evenly allocate data to the servers that make up the distributed database. By using these technologies in the collaborative field trials with Toyota Motor Corporation, we were able to achieve real-time vehicle data storage and search.

5. Future outlook

NTT Human Informatics Laboratories is developing a technology called Digital Twin Computing [10], which copies all relevant objects in the real world into a virtual space in real time. To enable this, it is essential to store and update, in real time, data that are transmitted simultaneously from all moving objects, including not only humans and vehicles but also drones, ships, and satellites, and to search an area of any shape, such as a city (buildings, roads, etc.) and environments (rivers, etc.), for the necessary data. Therefore, we are studying the extension of indexing in the height dimension, which is currently not supported by Axispot™, as well as the searching of 3D polygons.



API: application programming interface

Fig. 6. Axispot™ architecture.

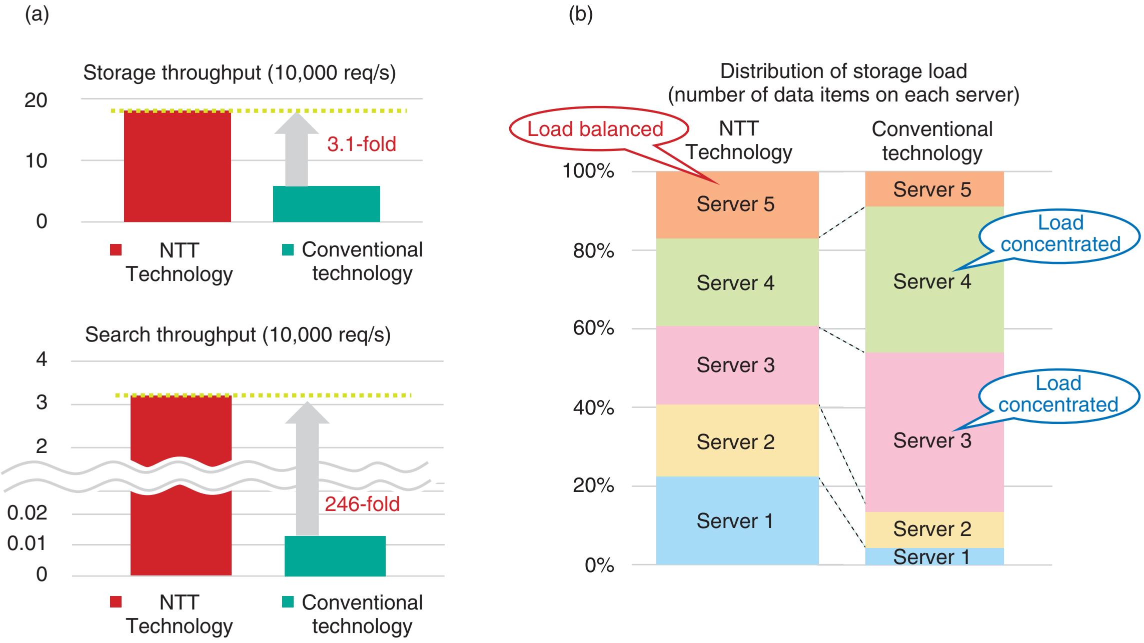


Fig. 7. Evaluation results.

References

[1] A. Isomura, I. Naito, Y. Iida, and T. Nakamura, “Axispot: A Distributed Spatiotemporal Data Management System for Digital Twins of Moving Objects,” IEEE Software, Vol. 39, No. 2, pp. 33–38, 2022. https://doi.org/10.1109/MS.2021.3132899

[2] A. Isomura, K. Miyahara, I. Naito, and M. Hanadate, “Real-time Spatial-temporal Database for Geographic Polygon Data Using HD-map,” Proc. of the 4th International Conference on Geoinformatics and Data Analysis (ICGDA 2021), pp. 7–13, April 2021. https://doi.org/10.1145/3465222.3465230

[3] T. Abraham and J. F. Roddick, “Survey of Spatio-temporal Databases,” GeoInformatica, Vol. 3, pp. 61–99, 1999. https://doi.org/10.1023/A:1009800916313

[4] N. Pant, M. Fouladgar, R. Elmasri, and K. Jitkajornwanich, “A Survey of Spatio-temporal Database Research,” Lecture Notes in Computer Science (LNCS), Vol. 10752, pp. 115–126, 2018.

[5] I. E. Sutherland, R. F. Sproull, and R. A. Schumacker, “A Characterization of Ten Hidden-surface Algorithms,” ACM Computing Surveys, Vol. 6, No. 1, pp. 1–55, 1974.

[6] M. Visvalingam and J. D. Whyatt, “The Douglas-Peucker Algorithm for Line Simplification: Re-evaluation through Visualization,” Computer Graphics Forum, Vol. 9, No. 3, pp. 213–225, 1990. https://doi.org/10.1111/j.1467-8659.1990.tb00398.x

[7] Redis, https://redis.io/

[8] Apache Ignite, https://ignite.apache.org/

[9] PostgreSQL, https://www.postgresql.org/

[10] Press release issued by NTT, “NTT Announces New R&D Projects of Digital Twin Computing,” Nov. 13, 2020. https://group.ntt/en/newsrelease/2020/11/13/201113c.html

Atsushi Isomura

Researcher, NTT Digital Twin Computing Research Center.

He received a Master’s degree in information science and technology from Aichi-Prefectural University in 2016. He joined NTT Software Innovation Center in 2016 and helped develop Axispot™, a distributed spatiotemporal database-management technology. Currently, he is working on digital-twin data-management technology at Digital Twin Computing Research Center. He received the Institute of Electronics, Information and Communication Engineers (IEICE) Tokai Section Student Award in 2015 and the best poster award in 12th International Conference on Ubiquitous Healthcare. He presented at RedisConf19 held in San Francisco.

Naoko Shigematsu

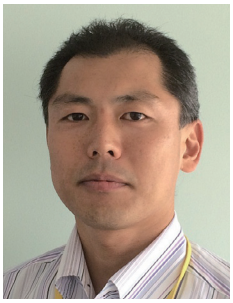
Research Engineer, NTT Digital Twin Computing Research Center.

She received a B.S. and M.S. in geophysics from Tohoku University, Miyagi, in 1993 and 1995. In 1995, she joined NTT Telecommunication Networks Laboratories and in 1999 moved on to NTT EAST R&D Center, where she engaged in the research of asynchronous transfer mode network operation systems. In 2000, she moved on to NTT Information Sharing Platform Laboratories, where she engaged in research of storage area networks. Since 2020, she has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Isoo Ueno

Research Engineer, NTT Digital Twin Computing Research Center.

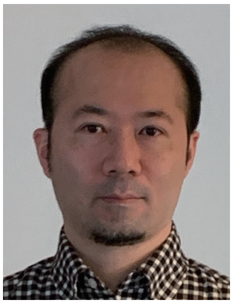
He received a B.S. and M.S. in earth science from Kobe University, Hyogo, in 1990 and 1992. He joined NTT in 1992 and engaged in basic research on complex systems. In 1998, he launched internet businesses in Hong Kong and UK. He also managed the development projects of OCN web hosting, Biz authentication, and the global billing system of NTT Communications. In 2021, he helped develop the polygon-based data-aggregation technology of Axispot™. He is currently working on Digital Twin Computing.



Nobuhiro Oki

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received a B.E. and M.E. in chemistry from Tokyo Institute of Technology in 1996 and 1996. He joined NTT in 1996 and launched the Internet telegram system (D-Mail). In 2008, he joined NTT Communications and launched various cloud services, including Smart Data Platform (formerly called Enterprise Cloud). In 2018, he joined NTT laboratories and helped develop Axispot™. He is currently working on digital-twin data-management technology.



Yutaka Arakawa

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received a B.E. and M.E. in electronic engineering from the University of Tokyo in 2001 and 2003 and Ph.D. in engineering from the University of Electro-Communications, Tokyo, in 2014. He joined NTT in 2003. His current research interests include distributed systems and data structure design. He is a member of IEICE.

Selective Vehicle-data-collection Algorithm

Masaru Takagi, Kazuya Matsuo, Ryota Nakada, and Koya Mori

Abstract

In an obstacle-detection use case, in which a monitoring system keeps track of an obstacle on the road, the system needs to continuously collect the latest images of the obstacle captured using onboard cameras. Although image-recognition technology can be used to accurately select relevant images (images that capture the obstacle in question), computational resources available in a vehicle are too limited to execute this task. In addition, transferring all images via a mobile network to the cloud incurs considerable communication costs. To solve these problems, we devised a technology that estimates the range of the area that can be captured with each camera (hereafter, visible range) and selectively collects only the relevant images. The visible range is estimated on the basis of meta-information, such as vehicle position, direction of movement, and camera angle of view.

Keywords: connected vehicle, data collection, priority control

1. Potential of connected vehicles

Connected vehicles are expected to become key players in using information and communication technology in road transportation. They also hold great potential as a sensing platform. Since connected vehicles are equipped with various high-end sensors and communication modules and move around a city with virtually no risk of running out of battery, they are ideal as mobile sensor nodes. If camera images and LiDAR (light detection and ranging)* point-cloud data from tens of millions of connected vehicles could be collected, it would be possible to continuously scan an entire city and build a digital twin of that city.

2. Assumed use case

Unfortunately, the total amount of sensor data generated by tens of millions of connected vehicles ranges from 10 to 100 Tbit/s. This is too enormous for communication networks, computers, and storage units to handle. This means that it is impractical to collect all available sensor data. It should be noted that sensor data include not only data that should be collected immediately but also data for which some collection delay is tolerable, data for which periodic collection is sufficient, and data that are of no value. Therefore, it is important to selectively collect important sensor data on a priority basis. By adjusting the pace and timing of sensor-data collection in accordance with the amount of load on the communication network and computers, it is also possible to equalize the load fluctuation over time, thus improve facility-utilization efficiency.

The collaboration projects between Toyota Motor Corporation and the NTT Group include several use cases that require the selective collection of sensor data. This article focuses on the obstacle-detection use case and introduces the technical challenges, implementation details, and our efforts to improve the performance of selective collection. In this use case, obstacles on the road are assumed to be moved by the wind or removed by the road administrator. Therefore,

---

* LiDAR: A type of remote sensing technology using light. It measures the distance to an object on the basis of the reflection time of a laser light emitted in all directions.

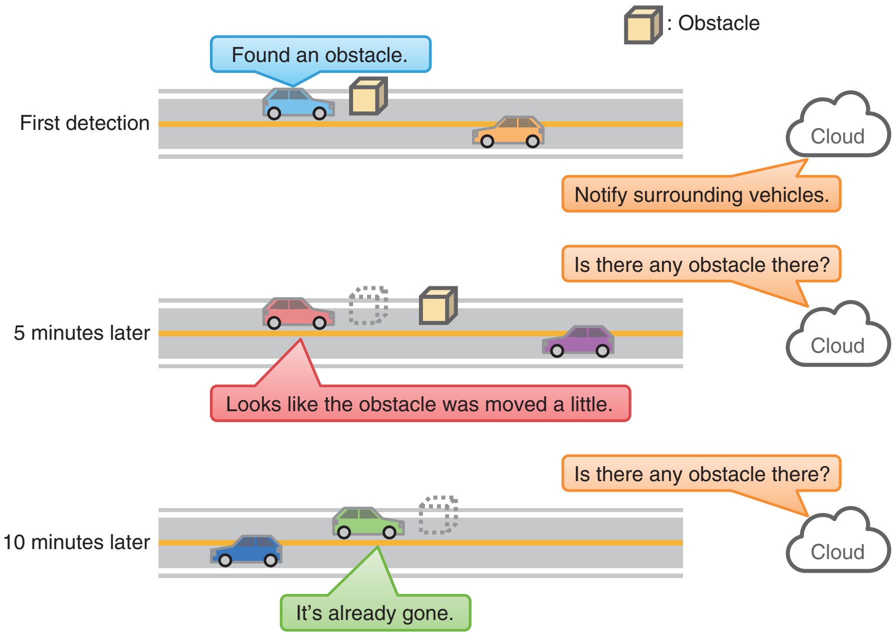


Fig. 1. Overview of the obstacle-detection use case.

the monitoring system keeps track of an obstacle on the road by collecting images from vehicles passing in the vicinity of the obstacle (**Fig. 1**).

3. Approach based on image recognition

At first glance, the use of image recognition seems to be the obvious approach to select images that capture the obstacle to be monitored. This approach can be implemented in two different configurations, but both present significant problems.

Configuration 1: Image recognition is executed within connected vehicles

The simplest configuration is to mount a computer on a connected vehicle to execute image recognition. However, this configuration has three problems unique to vehicles.

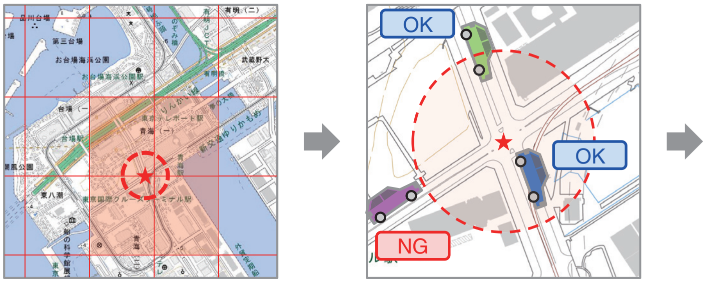
First, the temperature inside a vehicle under the burning sun in mid-summer can reach close to 80°C [1]. At such a high temperature, it is not easy to keep the computer operating stably. Second, it is said that the average usage rate of passenger cars is only 5%. If the computer stays idle most of the time, having a computer in a passenger car is not cost-effective. Third, the product life cycles of vehicles and computers differ greatly. The computer will reach the end of its life cycle or become obsolete in terms of performance much earlier than the vehicle on which it is mounted. Considering these problems, it is not realistic to install a computer capable of image recognition in a connected vehicle.

Configuration 2: Image recognition is executed in the datacenter

If there is a connected-vehicle platform in place, all images can be transferred to the datacenter so that image recognition can be executed there. Although this configuration prevents the three problems mentioned above, the cost of image transfer over the mobile network is very high. The load on the mobile network is also enormous because images are sent from tens of millions of vehicles simultaneously. Even if sensor-data traffic from a connected vehicle is just 1 Mbit/s, the total traffic from 10 million vehicles can reach 10 Tbit/s. This is 15 times the current volume of mobile communication traffic in Japan [2].

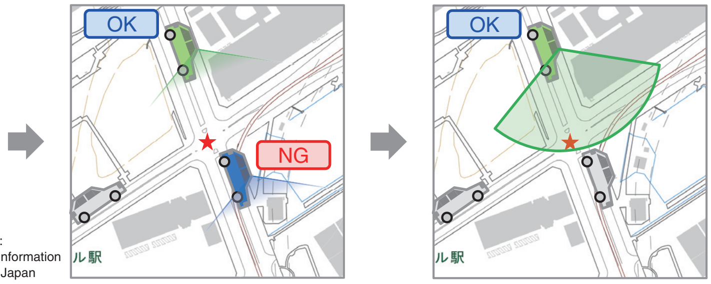
4. Approach based on location information

Faced with the problems mentioned above, we re-examined the requirements for the obstacle-detection use case. There is no need to detect an obstacle with high precision during image selection because the collected images are analyzed in detail at a later stage of processing. The number of images to be analyzed should be kept as small as possible because the processing for image analysis requires a large amount of



(1)(2) Maintain and search the vehicle-information database

(3) Determine the distance



(4) Determine the angle of view

(5) Detect shielding vehicles

Map source:
Geospatial Information
Authority of Japan

Fig. 2.   Process of image selection based on meta-information.

computation. Since the location of the obstacle to be monitored is already known, there is also no reason not to use this information as a clue.

In light of the above, we devised a technology that selectively collects images without relying on image recognition. Specifically, it estimates the visible range of the onboard camera of each vehicle on the basis of meta-information, such as the locations of surrounding vehicles, direction of the vehicle's movement, the vehicle's size, the angle of view of its camera, and the camera's resolution. Images that are likely to have captured the obstacle are then selected on the basis of the positional relationship between the obstacle and each vehicle. This image-selection process consists of the following five steps (**Fig. 2**).

Process 1: Maintain a vehicle-information database

Vehicle meta-information, which provides important clues for image selection, is routinely collected from connected vehicles and stored in the real-time spatiotemporal database Axispot™.

Process 2: Search the vehicle-information database

When a request for obstacle monitoring arrives, the system searches the vehicle-information database using the spatiotemporal index associated with the obstacle location and the current time and selects the search hits as candidate vehicles.

Process 3: Determine the distance

Next, the distance between each candidate vehicle and the obstacle is calculated, and any vehicle with a distance from the obstacle greater than the visible range of its onboard camera, which is calculated from the resolution of the onboard camera, is eliminated from the list of candidates.

Process 4: Determine the angle of view

Next, the direction of the obstacle from each candidate vehicle is calculated, and any vehicle that is estimated not to have captured the obstacle within the angle of view of its onboard camera is eliminated from the list of candidates.

Process 5: Detect shielding vehicles

Finally, the positional relationship between surrounding vehicles and the obstacle on the road is examined, and any vehicle that is found to be unable to detect the obstacle because its view of the obstacle is blocked (shielded) by other vehicles is eliminated from the list of candidates.

5. Reducing complexity for shielding vehicle detection

Of these five steps, the detection of shielding vehicles demands the most complex computation and

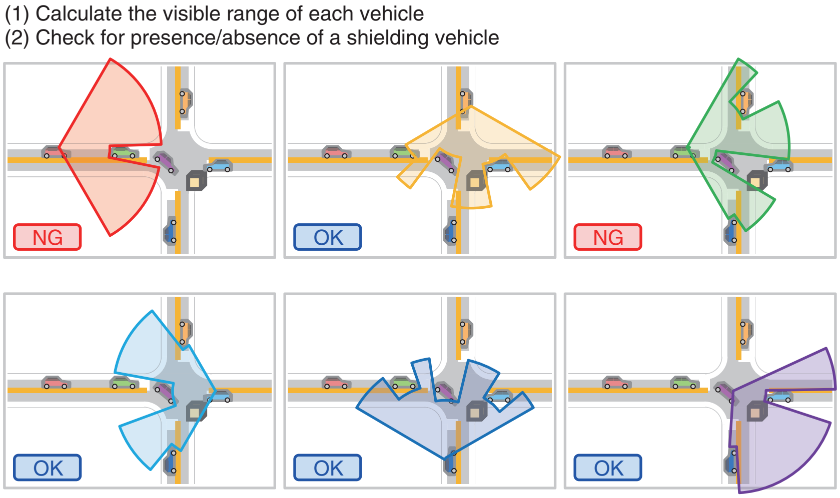


Fig. 3. Example of checking for presence/absence of a shielding vehicle (before algorithm improvement).

accounts for the bulk of the computation time. Therefore, during the field trial period, we attempted to substantially improve the algorithm for shielding vehicle detection. This section presents the algorithm before and after this improvement, in this order.

The most intuitive method of detecting a shielding vehicle is to calculate the visible range of each candidate vehicle and determine whether there is any vehicle within this range that blocks the view of the obstacle. For example, suppose there are six vehicles in the vicinity of an obstacle, as shown in **Fig. 3**. A fan-shaped field of view is drawn for each of the six vehicles on the basis of the angle of view and visible range of its camera. The system then checks whether a vehicle's view of the obstacle is blocked (shielded) by any of the other five vehicles and eliminates any vehicle that is unable to detect the obstacle. Since this method requires a brute-force examination of any shielding vehicle, the computational complexity is $O(n^2)$, where n is the number of vehicles. This means that, as the number of vehicles in the vicinity of the obstacle increases, the processing time increases rapidly. Therefore, in situations where many vehicles are densely packed around the obstacle due to traffic congestion, the processing time can become very long.

We, therefore, took on the challenge of improving the algorithm to reduce the amount of computation. The key point of this improvement effort was to abandon the conventional approach of calculating the visible range of each onboard camera. We reversed the thinking: from examining whether a vehicle can detect the obstacle to examining whether the obstacle can detect the vehicle (i.e., whether the view of the vehicle from the obstacle is blocked by another vehicle). The system calculates the visible range of the obstacle instead of the visible range of the vehicle to determine whether any vehicle blocks the view of the obstacle from each vehicle (**Fig. 4**). With this method, the system examines each vehicle, one by one, to check whether it blocks the view of other vehicles. This eliminates the need for a brute-force calculation and reduces the computational complexity to O(n). Also, by approximating the fields of view as a set of fan shapes, the process of determining whether a vehicle blocks the view of others can be implemented in a very simple manner.

6. Performance evaluation

We evaluated the processing time before and after the algorithm improvement for various vehicle location patterns (**Fig. 5**). It was assumed that the visible range of the onboard cameras was 100 m. The number of vehicles posited to be present within a 100-m radius of the obstacle was varied. To evaluate the worst case, it was important to set a sufficiently high upper limit for the number of vehicles. We set the upper limit at 2000 vehicles, assuming that the vehicles were packed as densely as in a parking lot. Java was used as the language for implementing the selection algorithm. A geometric calculation library called JTS Topology Suite [3] was used with the algorithm before the improvement, but not after the improvement.

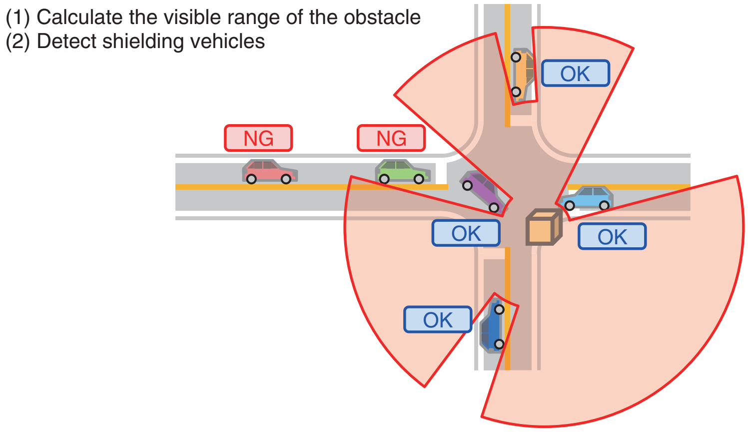
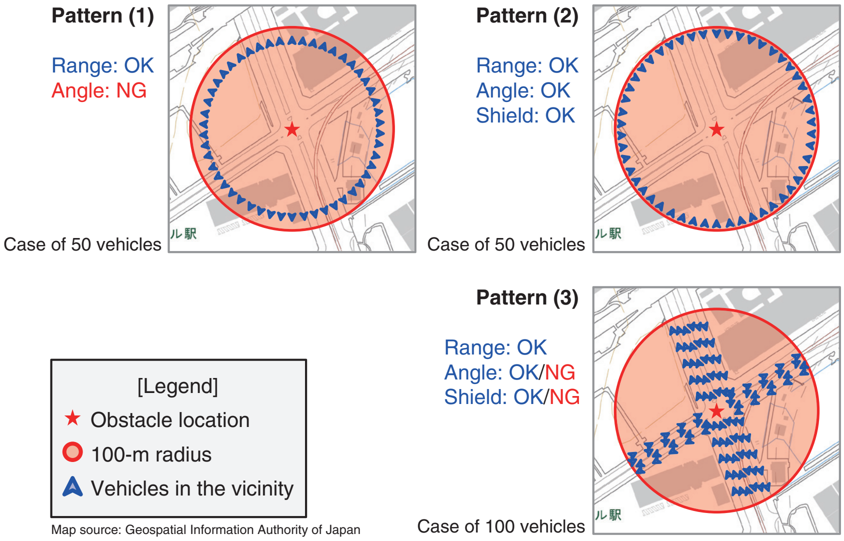


Fig. 4. Example of shielding determination (after algorithm improvement).



Map source: Geospatial Information Authority of Japan

Fig. 5. Vehicle-location patterns in performance evaluation.

Figure 6 shows the processing time for each vehicle-location pattern with varying the number of vehicles. Before the algorithm improvement, the processing time for detecting shielding vehicles increased rapidly in proportion to the square of the number of vehicles. After the algorithm improvement, the processing time was proportional to the number of vehicles. In the worst-case scenario with 2000 vehicles, the processing time for detecting shielding vehicles was 25 ms, which is 1/60 of the processing time before the algorithm improvement.

7. Future outlook

In these field trials, we optimized the shielding-vehicle-detection algorithm for the requirements of the obstacle-detection use case and achieved a processing speed 60 times faster than that for when the algorithm was implemented in a conventional manner using a geometric calculation library. In the future, we will investigate the possibility of using our study results for applications other than connected vehicles.

For example, when creating a digital twin of a city, it is necessary to ensure that it reflects the latest state of the city by periodically re-sensing every area.

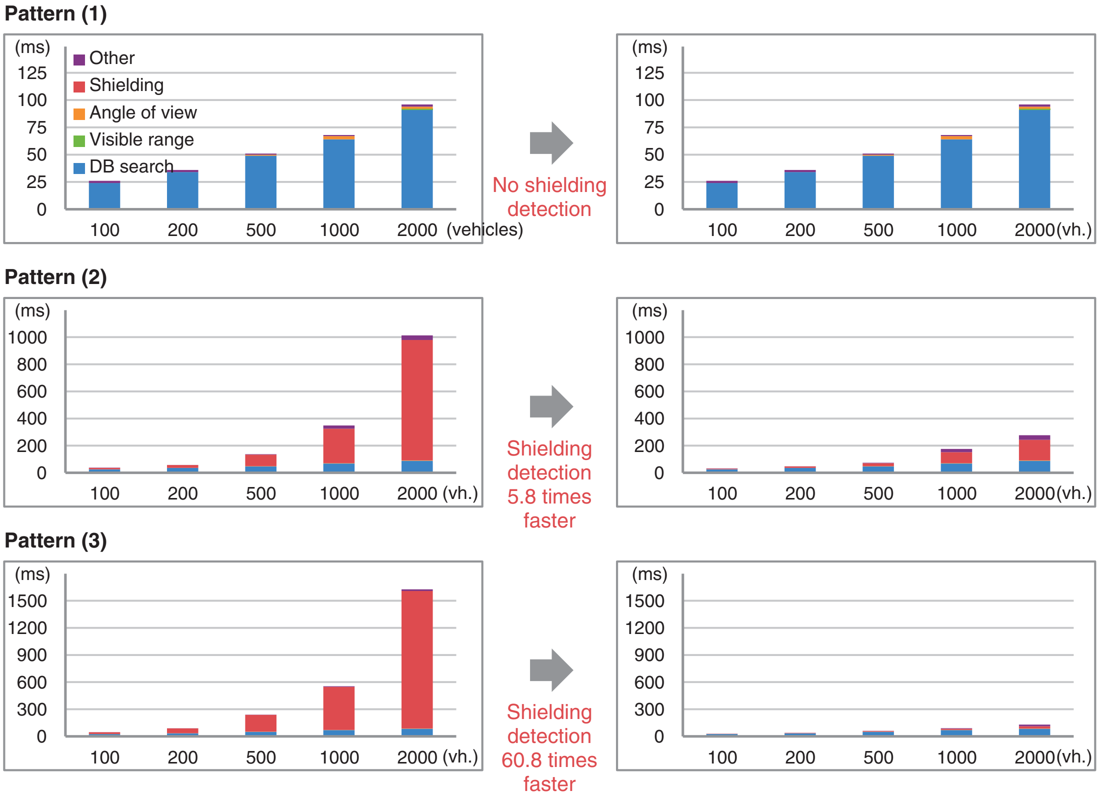


Fig. 6. Vehicle-location patterns vs. processing times.

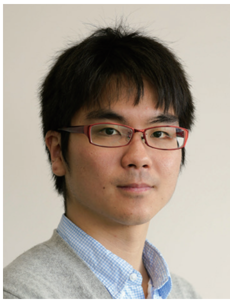
Some parts of the cityscape change more frequently than others. Therefore, it is more efficient to systematically re-sense only key areas frequently than to uniformly re-sense the entire city. In such situations, the algorithm presented in this article could be used to efficiently select data to be collected. We will also study application-related issues, such as issues that may arise when the monitoring targets are widely spread or when position information contains errors.

References

[1] Japan Automobile Federation, “Temperature Inside a Vehicle in Midsummer” (in Japanese), https://jaf.or.jp/common/safety-drive/car-learning/user-test/temperature/summer

[2] Ministry of Internal Affairs and Communications, “Aggregation and Provisional Calculation of Internet Traffic in Japan—Announcement of Aggregate Results as of November 2021,” Feb. 2022. https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pressrelease/2022/2/04_02.html

[3] The JTS Topology Suite, https://locationtech.github.io/jts/



Masaru Takagi

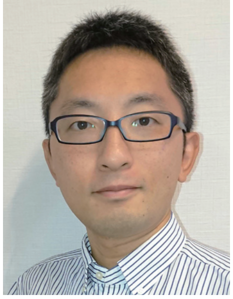
Researcher, NTT Digital Twin Computing Research Center.

He received a B.E. in engineering and M.E. and Ph.D. in information science and technology from the University of Tokyo in 2013, 2015, and 2018. He joined NTT in 2018 and was engaged in a collaboration project with Toyota Motor Corporation. He is currently studying multi-agent mobility simulation as part of Digital Twin Computing. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and Information Processing Society of Japan (IPSJ).

Kazuya Matsuo

Researcher, NTT Digital Twin Computing Research Center.

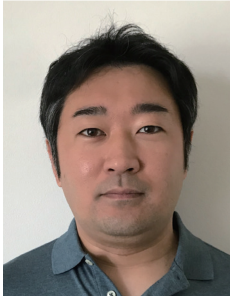
He received a Ph.D. in information science from Osaka University in 2017 and joined NTT the same year. While at university (2012–2017), he studied the collection of data in mobile sensor networks. He studied modeling of vertically distributed applications in NTT until 2019. Since 2018, he has been involved in collaborative work between Toyota and NTT. Since 2020, he has also been working on a Digital Twin Computing project. He received a Young Researcher's Award from IEICE in 2018. He is a member of IPSJ.



Ryota Nakada

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received a B.E. and M.E. in mathematics from Tokyo University of Science in 2008 and 2010. He joined NTT Network Innovation Laboratories in 2010 and researched a high-resolution image-transport system. He moved on to NTT WEST in 2013, where he studied an indoor-location information system. He rejoined NTT Network Innovation Laboratories in 2016 and has been researching edge computing. Since 2020, he has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.



Koya Mori

Senior Research Engineer, NTT Digital Twin Computing Research Center.

Since joining NTT in 2004, he has been involved in the research and development of an Internet-of-Things application based on software technologies, such as OSGi, OpenStack, and Edge Computing. Since 2020, he has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Vertically Distributed Computing Technology

Kazuya Matsuo, Masaru Takagi, Ryota Nakada, and Koya Mori

Abstract

We developed a technology that quickly shares information found by a connected vehicle with other vehicles. For example, when a connected vehicle finds an obstacle on the road, the technology can transmit information about it quickly to other vehicles. This quick notification is achieved by offloading part of the processing of the collected obstacle information to network-edge nodes and transmitting interim results to vehicles near the obstacle. However, in urban areas, the delay in notification may become large because a large number of vehicles connect to a small number of network-edge nodes; thus, the volume of data received from these vehicles may overwhelm the computing resources of these nodes. To solve this problem, we also developed a technology that dynamically selects which computer will execute any particular notification processing on the basis of the states of the connected vehicles. Using this technology, obstacle information can be transmitted quickly even when a large number of vehicles are connected to a small number of network-edge nodes.

Keywords: distributed processing, edge computing, load balancing

1. Roles and challenges with an ICT platform for connected vehicles

The role of an information and communication technology (ICT) platform for connected vehicles is to collect videos captured with onboard cameras and CAN (controller area network) data from connected vehicles, analyze them, and send the analysis results to the other vehicles to provide these vehicles with more information than they can obtain from their sensors alone. This enables drivers to obtain information about their blind spots, preventing traffic accidents from occurring.

When connected vehicles are widely used, a large amount of data will be collected from a large number of such vehicles and may overwhelm the ICT platform. Even in such situations, the ICT platform must be able to quickly notify vehicles of urgent information, such as that relating to obstacles.

In the collaboration between the NTT Group and Toyota Motor Corporation, we conducted field trials for several use cases in which the ICT platform notifies relevant vehicles about the processing results for the data it has collected.

The vertically distributed computing introduced in this article consists of the *vertically distributed application architecture* and *technology for dynamically selecting processing nodes* we developed. The following sections describe the challenges each of these technologies addresses, solutions each provides, and verification results of the field trials.

2. Vertically distributed computing

2.1 Vertically distributed application architecture

(1) Obstacle detection and notification in the previous ICT platform

The previous ICT platform used in our collaboration with Toyota was built on the lambda architecture [1, 2], in which stream (real-time) data processing and batch data processing are executed in parallel, allowing both quick results from stream processing and detailed results from batch processing to be made available (**Fig. 1**). To be able to execute obstacle

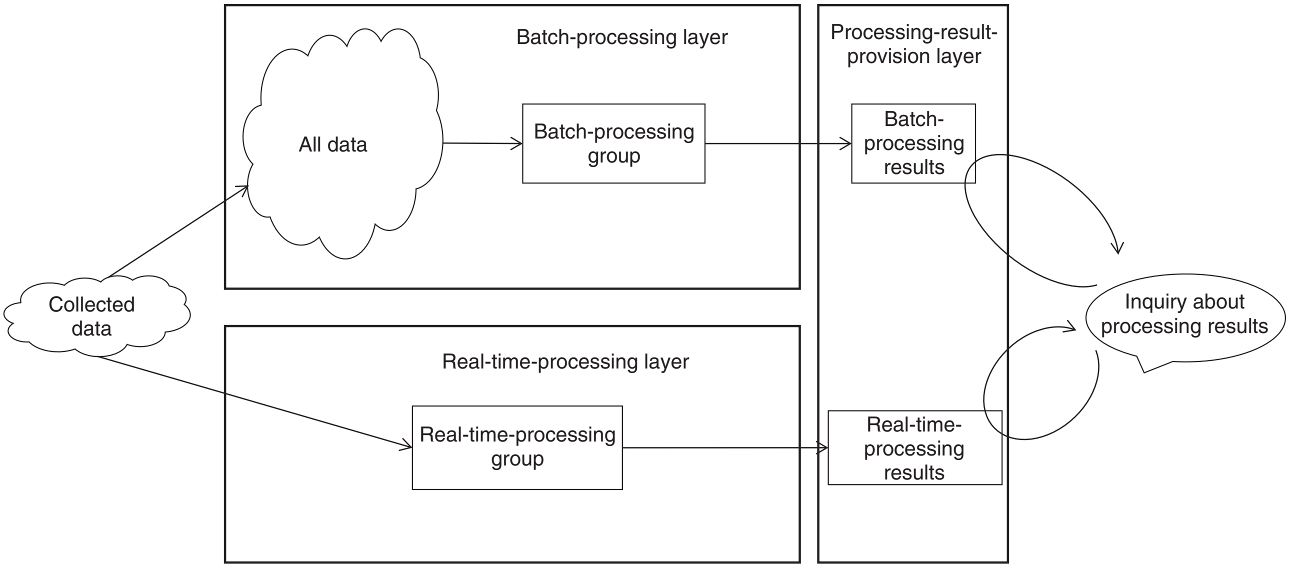


Fig. 1. Lambda architecture.

detection and notification within 7 s, the platform estimated the type and location of any obstacle from each frame (image) by stream processing. However, the lambda architecture does not take into account the requirements for processing time. Therefore, a system built on this architecture does not necessarily satisfy the requirement to execute obstacle detection and notification within 7 s. The actual processing time has been about 15 s.

(2) Vertically distributed application architecture

To solve this problem, we proposed an architecture that allows a processing-time threshold to be set in the lambda architecture. This new architecture is called a vertically distributed application architecture (**Fig. 2**). In this architecture, groups of processes are extracted from a series of processes in the order of their processing in such a way that the total processing time does not exceed the threshold. The results of these extracted processes can be obtained by the client. We revised the ICT platform to run on this architecture. In the revised ICT platform, object recognition is extracted from image-data processing, which consists of object recognition and object-location estimation. By doing so, connected vehicles can obtain information about the type of obstacle before the ICT platform completes the estimation of the obstacle location. To further increase the processing speed, the revised ICT platform offloads[*1] object recognition to network-edge nodes[*2], which are located closer to the relevant connected vehicles than the datacenter servers are. These nodes send obstacle information directly to the connected vehicles. In this architecture, obstacle detection and notification are executed in the following two stages (**Fig. 3**):

(i) First notification: Information about the type of obstacle and the rough obstacle location (actually, the location of the connected vehicle that took the image of the obstacle) is sent to connected vehicles near the obstacle.

(ii) Notification of details: The type of obstacle and more precise information about its location, as obtained from the obstacle location estimation, are sent to connected vehicles near the obstacle.

(3) Field trial

In a field trial, the vehicle that took the image of the obstacle and the vehicle that received the obstacle information were the same. We measured the time between when the onboard camera captured the obstacle and when the vehicle received the obstacle information from the ICT platform. The results are shown in **Table 1**. The first notification was sent within 7 s. As mentioned above, the first notification only contained information about the approximate location of the obstacle because the precise location had not yet been calculated at this point. Since this

---

[*1] Offloading: A mechanism by which the processing load of a computer is reduced by passing it to another computer.

[*2] Network-edge node: A set of computers that are located at NTT's base stations and central offices.

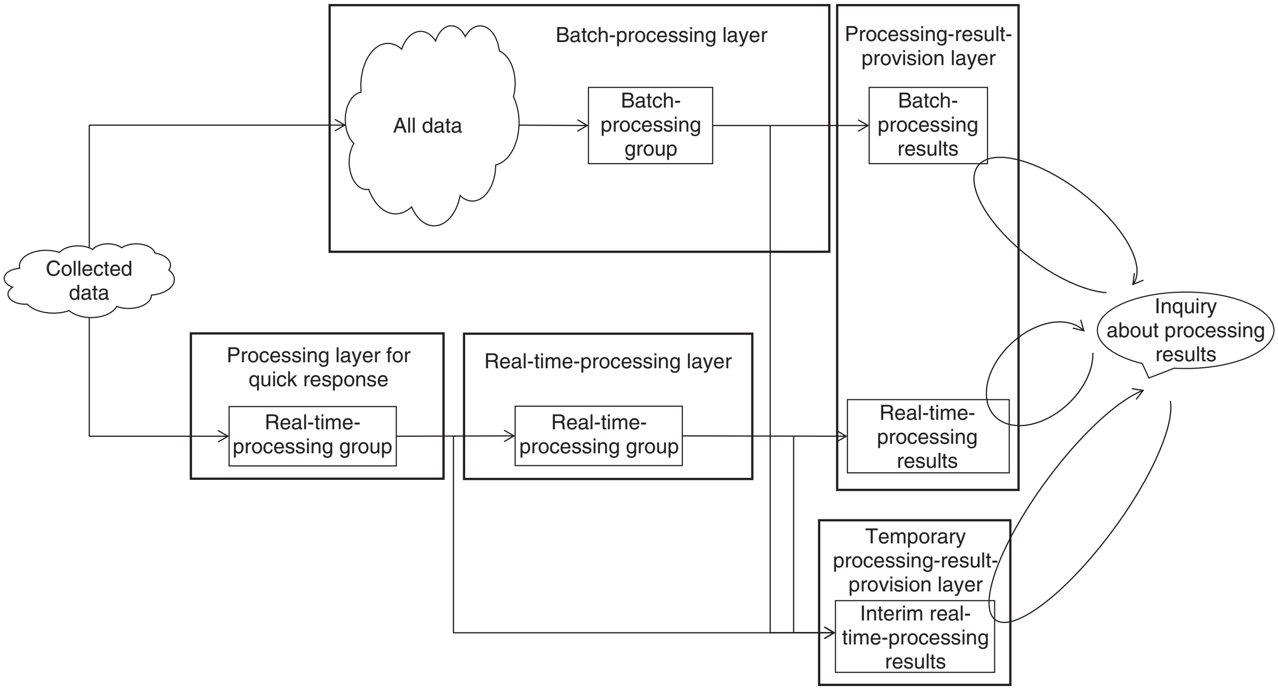


Fig. 2. Vertically distributed application architecture.

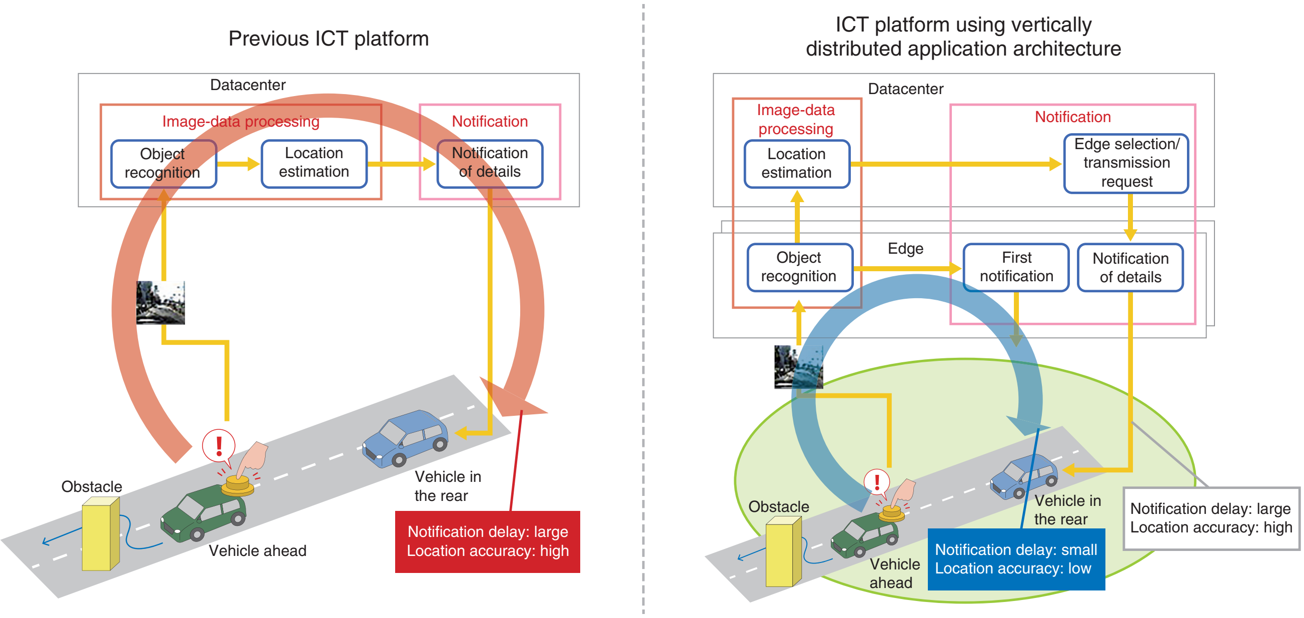


Fig. 3. A before-and-after comparison for the introduction of vertically distributed application architecture.

approximate location was the location of the vehicle that took the image of the obstacle, it was several to ten meters away from the real obstacle location. However, since vehicles can move more than 10 m/s, we believe that providing information about the approximate obstacle location more than 10 s earlier

Table 1. Results of the field trial of vertically distributed application architecture.

Processing time			
From obstacle detection to first notification		From obstacle detection to notification of details	
Average (s)	Fastest (s)	Average (s)	Fastest (s)
4.574	4.092	9.772	9.212

than had been possible, even at the cost of location accuracy, would be extremely valuable for assisting safe driving.

2.2 Dynamic selection of processing nodes

(1) Issues confronting the ICT platform after the adoption of the vertically distributed application architecture

The field trial showed that the vertically distributed application architecture made quick notification possible. However, the trial was conducted under conditions such that the computing resources of the network-edge nodes were not exhausted. When connected vehicles are widespread, a large number of them may be connected to a small number of network-edge nodes. The concentration of the processing load for object recognition on these edge nodes may overwhelm their computing resources, making it impossible to send notifications quickly.

One way to distribute a load that would otherwise concentrate on a small number of computers is to offload the processing load to the most appropriate computers on the basis of the acceptable processing delay and remaining computer resources. Many such technologies have been proposed (e.g., [3]). However, adopting only these conventional technologies cannot solve the above problem because vehicles can move more than 10 m/s, causing the surrounding conditions to change quickly from moment to moment, and the acceptable processing delay varies depending on these surrounding conditions. For example, if an obstacle is located on a blind curve, the risk of an accident is very high, and it is urgent to alert the vehicles near the obstacle. In contrast, if an obstacle is on a road with good visibility, the risk of an accident is low, and it is less urgent to send obstacle notifications to the vehicles near the obstacle. Conventional technologies do not take into account these client situations (i.e., the road situations mentioned above). Therefore, if the ICT platform's load balancing is based on conventional technologies, obstacle notification may not be sent quickly even in situations where quick notification is absolutely needed, or it may be sent quickly even when there is no urgent need for it to be.

(2) Overview of technology for dynamically selecting processing nodes

To address the issue mentioned above, we developed a technology for dynamically selecting processing nodes. It adds to the conventional load balancing technology the ability to take into account the surrounding conditions of the vehicles to be notified. This technology consists of the following three processes:

(i) Determine data-processing priority on the basis of the surrounding conditions of the vehicles to be notified (the part added in the proposed technology)
(ii) Select the computers to which the target processing is offloaded (conventional technology)
(iii) Execute the above offloading (conventional technology)

Various algorithms for Process (i) can be conceived for each use case. In the field trial, we adopted an algorithm designed for the obstacle detection and notification use case and verified its effectiveness.

(3) Application of the technology for dynamically selecting processing nodes to the ICT platform

In the field trial, we adopted an algorithm that takes into account the requirement that connected vehicles moving at high speed without being able to detect the obstacle should be urgently notified. Specifically, the platform searches for connected vehicles in the vicinity of the vehicle that transmitted the relevant onboard camera image. These are the targets for obstacle notification. The platform then determines whether it is urgent to notify these connected vehicles on the basis of the following two conditions (**Fig. 4**):

(i) Whether the average speed of the target connected vehicle exceeds the threshold.
(ii) Whether the obstacle is visible from the driver's seat (to be more precise, whether the driver can see the location of the connected vehicle that sent the relevant obstacle-containing image when the object is first detected, or whether the driver can see the estimated location of the obstacle when the obstacle has already been precisely detected and is being monitored).

To implement Condition (ii), the platform uses an algorithm similar to the selective vehicle-data-collection algorithm to determine whether the view of the obstacle is blocked. If quick notification is required,

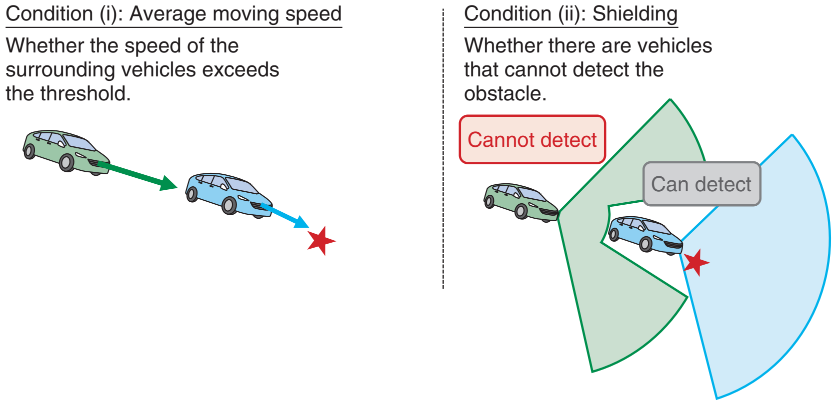


Fig. 4. Determining whether there are connected vehicles requiring quick notification.

Table 2. Setting of processing priority for images from onboard cameras.

(i) Average moving speed	(ii) Shielding	Priority
High	Yes	High
	No	Medium
Low	Yes	Medium
	No	Low

Table 3. Selection of the processing computer based on priority.

Priority	Processing computer
High	Nearby network-edge node with remaining resources
Medium	Nearby computer with remaining resources (network-edge node or datacenter server)
Low	Datacenter server

the processing priority is set to high. If quick notification is not required, the processing priority is set to low. For in-between cases, the processing priority is set to medium. This priority setting is summarized in **Table 2**.

The next step is to determine, based on the priority setting just described, which computer should execute the object recognition. As with the conventional technology, our technology for dynamically selecting processing nodes monitors the remaining resources of each computer in the system and selects the appropriate computer on the basis of the remaining resources of each computer and the processing priority, as shown in **Table 3**. Finally, the platform offloads the object-recognition process to the selected computer and the latter executes the process.

Figure 5 illustrates an example computer configuration in the ICT platform and how the computers to which the target process is offloaded are selected. We assume that the network-edge node receives the onboard camera videos in the sequence of videos 1 to 6 in the figure and that the object-recognition priority for each video is set as shown in the figure. Object recognition 4, which processes video 4, has low priority thus offloaded to a datacenter server. Object recognition 5 has medium priority. It is executed by the network-edge node because the above offloading has freed up its resources. This exhausts the computing resources at the network-edge node. Therefore, object recognition 6, which has medium priority, is offloaded to a datacenter server.

(4) Field trial

In the field trial, we examined how the number of onboard camera videos sent to the network-edge node affects the image-processing time. The results are shown in **Fig. 6**. The horizontal axis shows the number of videos received per second. The number of high-priority videos was varied between 1 and 3, and the number of medium-priority videos was varied between 1 and 9. The vertical axis shows the increase in processing time from a condition under which computing resources are sufficiently available. As can be seen from this figure, when no dynamic selection of processing nodes was used, the processing time at the network-edge node increased when more than three videos were received per second. In contrast, when dynamic selection of processing nodes

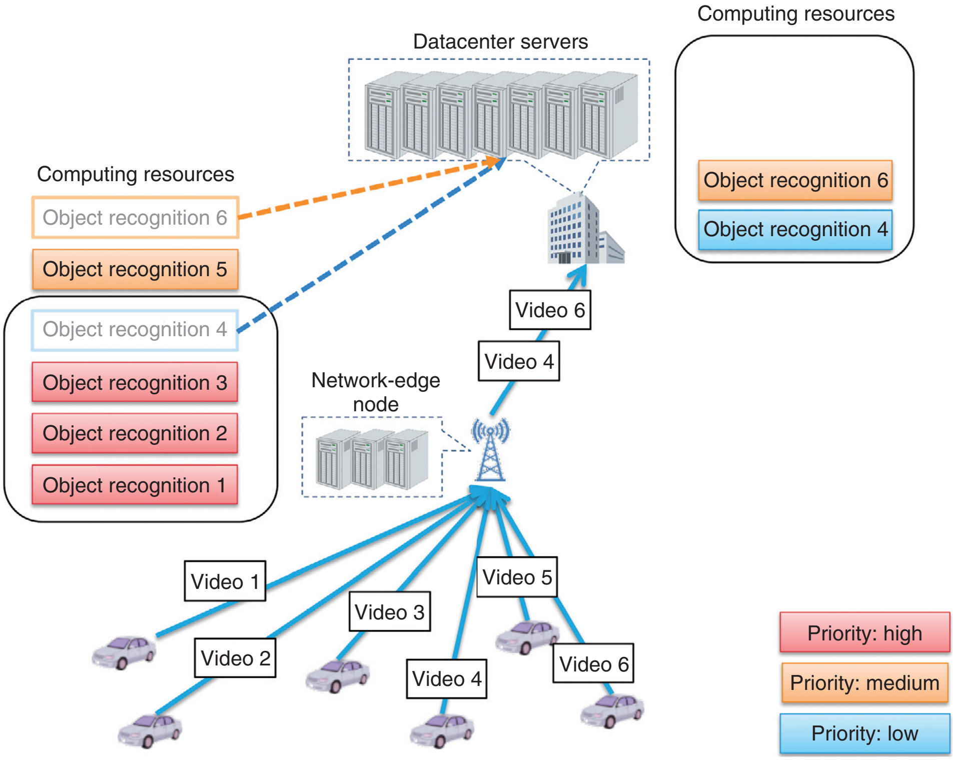
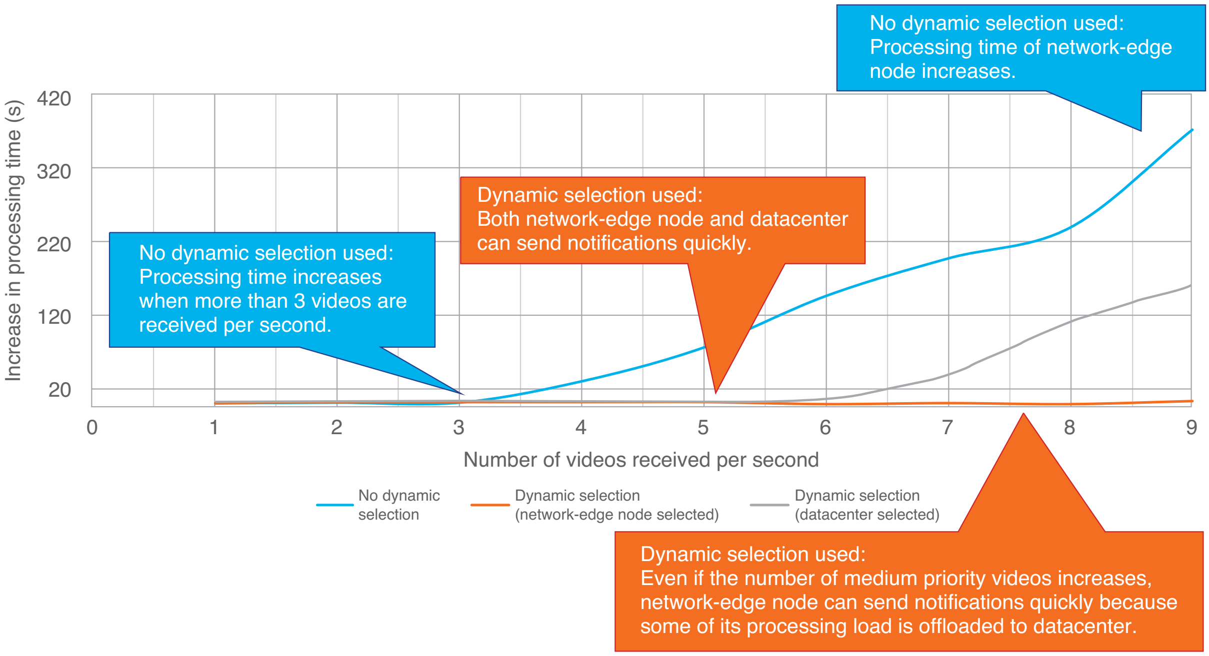


Fig. 5. Example of the dynamic selection of processing nodes.



No dynamic selection used:
Processing time of network-edge node increases.

Dynamic selection used:
Both network-edge node and datacenter can send notifications quickly.

No dynamic selection used:
Processing time increases when more than 3 videos are received per second.

Dynamic selection used:
Even if the number of medium priority videos increases, network-edge node can send notifications quickly because some of its processing load is offloaded to datacenter.

Fig. 6. Results of the field trial of dynamically selecting processing nodes.

was used, the processing time at the network-edge node did not increase until nine[*3] videos were received per second. This was because the network-edge node processed only high-priority processes and other processes were offloaded to the datacenter. Because of this offloading, the processing time at the datacenter began to increase when six videos were received per second.

3. Future outlook

Since computer technology is advancing daily, the computers in vehicles have become powerful enough to process the video from their onboard cameras. However, it is not assumed with the current vertically distributed computing technology that connected vehicles can process onboard camera videos. To make more effective use of the computing resources at both network-edge nodes and datacenters, we will study the extended use of the computing resources in connected vehicles.

In the obstacle detection and notification use case, an additional type of processing is required: when the platform detects an obstacle in the image from a connected vehicle, it needs to determine whether the obstacle is being identified for the first time or has already been found in the image from another vehicle. This is because, if the obstacle has already been found, the connected vehicles near the obstacle will have already been notified of the obstacle, and so there is less urgency. We are also studying technology for determining whether the objects detected by multiple sensors are identical [4].

References

[1] N. Marz, "Big Data Lambda Architecture," 2012. http://www.databasetube.com/database/big-data-lambda-architecture/

[2] K. Dutta and M. Jayapal, "Big Data Analytics for Real Time Systems," Big Data Analytics Seminar, pp. 1–13, 2015.

[3] Y. Dai, D. Xu, S. Maharjan, and Y. Zhang, "Joint Load Balancing and Offloading in Vehicular Edge Computing and Networks," IEEE Internet of Things Journal, Vol. 6, No. 3, pp. 4377–4387, June 2019.

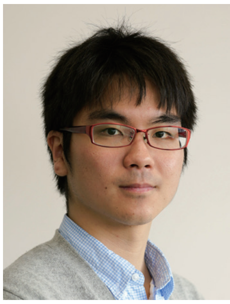
[4] K. Matsuo, M. Takaki, R. Nakata, and K. Mori, "A Study on a Dynamic Object Detection and Identification System That Uses Point Cloud Alignment for Creating a High-precision Dynamic Map," SIG Technical Reports on Human-Computer Interaction, Vol. 2021, No. 6, pp. 1–8, 2021 (in Japanese).

[*3] This load is very small for a world in which connected vehicles are widely used. This small value was derived because the field trials were intended to confirm the effectiveness of each technology studied in this collaboration, thus were conducted using the minimum computer configuration. When the ICT platform is built with a large-scale computer configuration to confirm practical feasibility in the future, a larger load will be applied when verifying the effectiveness of each technology.

Kazuya Matsuo

Researcher, NTT Digital Twin Computing Research Center.

He received a Ph.D. in information science from Osaka University in 2017 and joined NTT the same year. While at university (2012–2017), he studied the collection of data in mobile sensor networks. He studied modeling of vertically distributed applications in NTT until 2019. Since 2018, he has been involved in collaborative work between Toyota and NTT. Since 2020, he has also been working on a Digital Twin Computing project. He received a Young Researcher's Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2018. He is a member of the Information Processing Society of Japan (IPSJ).



Masaru Takagi

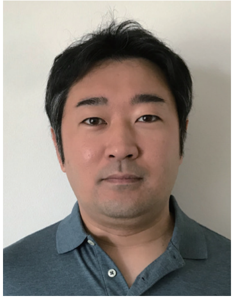
Researcher, NTT Digital Twin Computing Research Center.

He received a B.E. in engineering and M.E. and Ph.D. in information science and technology from the University of Tokyo in 2013, 2015, and 2018. He joined NTT in 2018 and was engaged in a collaboration project with Toyota Motor Corporation. He is currently studying multi-agent mobility simulation as part of Digital Twin Computing. He is a member of IEICE and IPSJ.

Ryota Nakada

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received a B.E. and M.E. in mathematics from Tokyo University of Science in 2008 and 2010. He joined NTT Network Innovation Laboratories in 2010 and researched a high-resolution image-transport system. He moved on to NTT WEST in 2013, where he studied an indoor-location information system. He rejoined NTT Network Innovation Laboratories in 2016 and has been researching edge computing. Since 2020, he has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.



Koya Mori

Senior Research Engineer, NTT Digital Twin Computing Research Center.

Since joining NTT in 2004, he has been involved in the research and development of an Internet-of-Things application based on software technologies, such as OSGi, OpenStack, and Edge Computing. Since 2020, he has been in charge of research for Digital Twin Computing at NTT Digital Twin Computing Research Center.

Lane-specific Traffic-jam-detection Technology

Kouhei Mori, Yuki Yokohata, Aki Hayashi, Takahiro Hata, and Masato Kamiya

Abstract

Dynamic maps are being constructed to support automated driving and advanced navigation. They combine high-precision map information with traffic-related information, such as that about traffic controls and jams. We at NTT believe that detecting lane-specific traffic jams, such as those caused by a queue of vehicles waiting to enter the parking lot of a commercial facility or by vehicles parked on the street, and providing information on these jams will enable unprecedentedly advanced navigation. This article describes a technology for detecting lane-specific traffic jams on the basis of information that can be collected from connected vehicles.

Keywords: lane-specific traffic jam, connected vehicle, traffic flow optimization

1. Introduction

Dynamic maps [1] are being constructed to support automated driving and advanced navigation. They combine high-precision map information with traffic-related information such as that about traffic controls and jams. Such traffic-related information is mapped on a dynamic map in real time, enabling map systems to detect road-traffic conditions and provide services, such as navigation, to optimize traffic flow. To construct a dynamic map, an approach is needed to collect, in real time, traffic-related information on roads across the country. Examples of such an approach are millimeter-wave radar [2] and traffic counters [3], which can be used to measure the volume of the flow of passing vehicles at fixed points, thus detect traffic jams. However, they must be installed at all locations where traffic jams are expected, so cannot cover a wide area. Another approach [4] involves using the smartphones of drivers to measure their vehicle speeds to detect traffic jams. This approach can cover a wide area but does not provide lane-specific resolution. In real-world traffic, lane-specific queues occur at various locations such as at the entrance to the parking lot of a commercial facility, at the exits of highways, or in turn lanes at intersections. When drivers who are new to an area encounter such a queue, they find it difficult to judge whether they need to stay in the queue to reach their destination. They may wait in a queue unnecessarily or avoid the queue when they actually need to join it, with the result that they must take a roundabout route to reach their destinations.

If lane-specific queues (traffic jams) and specific sections within a queue can be detected, it will be possible to provide unprecedented, advanced driving assistance, such as lane-specific navigation, that determines, on the basis of the driver's destination, whether the driver should join or avoid a queue in a particular lane, enabling drivers to reach their destinations without queuing up unnecessarily. This article describes a technology that detects lane-specific traffic jams on the basis of video data and vehicle speeds obtained from connected vehicles [5].

2. Lane-specific traffic-jam-detection technology

2.1 Definition

Taking a cue from the definition of a traffic jam applied to current services on general roads [6], we defined a lane-specific traffic jam as a line of traffic moving at 10 km/h or less. We also defined the

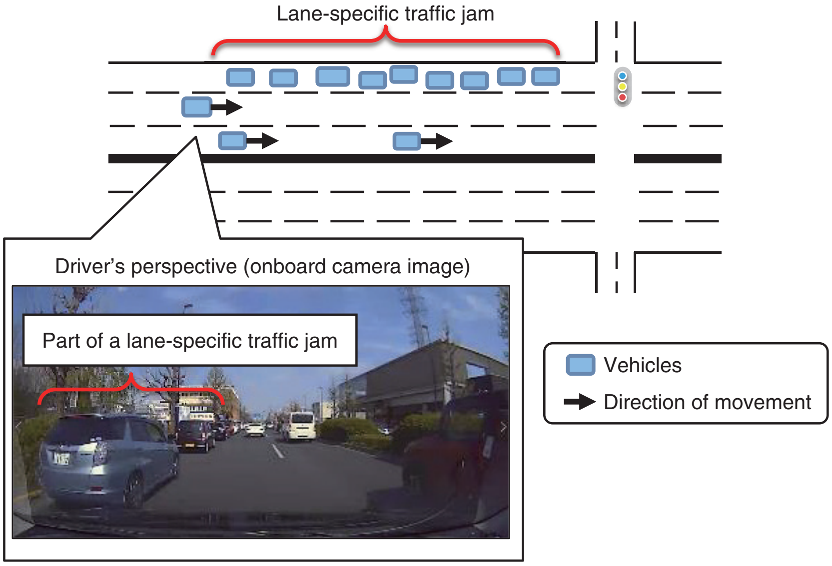


Fig. 1. Driver's perspective when encountering a lane-specific traffic jam.

minimum length of a traffic jam as 42 m, which is 1/10 the length defined in current services. In this article, the connected vehicle from which data are collected is referred to as an *observation vehicle*, and vehicles other than the observation vehicle that appear in images from the onboard camera mounted on the observation vehicle are referred to as *other vehicles*.

2.2 Processing overview

Figure 1 illustrates road conditions and an onboard camera image when a lane-specific traffic jam is encountered. From the onboard camera video taken with the observation vehicle, our technology counts other vehicles overtaken by the observation vehicle and that of other vehicles that passed in the opposite lane. The density of other vehicles can then be calculated from the speed of the observation vehicle and number of vehicles counted in each lane. A lane-specific traffic jam can be detected from this density calculation. When a lane-specific traffic jam is detected, the timeframe used to evaluate the traffic jam is shifted, and the evaluation is repeated to estimate the section where the lane-specific traffic jam is located and identify the vehicles at the front and back of the queue.

2.3 Processing flow

The front camera of the observation vehicle takes a video of the direction in which the vehicle is moving. From that video, the following processes are executed, as shown in **Fig. 2**.

(1) Detecting vehicles

The collected video is divided into frames, and an object-detection technology [7] is used to detect other vehicles in each frame. A detected vehicle is represented as a rectangle (detection rectangle) and labeled as a car, truck, or bus on the basis of the vehicle type (**Fig. 3**).

(2) Tracking

Since the relative positions between the observation vehicle and other vehicles change over time, the movements of other vehicles between frames are tracked and numbered. Next, a number is assigned to each of the other vehicles in each frame. The degree of similarity between the detected rectangle X in the current frame N and detected rectangle Y in the f-th frame back from N (i.e., Frame $N - f$) is evaluated using a similarity measure between two sets. This measure is called the intersection over union (IoU):

$$IoU = \frac{|X \cap Y|}{|X \cup Y|}. \qquad (1)$$

A pair of rectangles with the largest IoU, i.e., the rectangles with the most common pixels, are judged to be the same vehicle. This makes it possible to track the trajectory of this vehicle.

(3) Counting overtaken vehicles

The same reference lines are drawn on each frame (**Fig. 4**). The trajectory of the other vehicle obtained

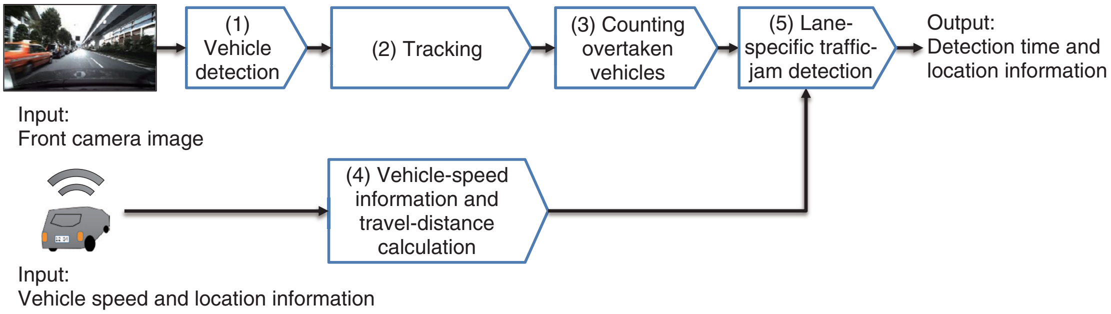


Fig. 2. Processing flow.

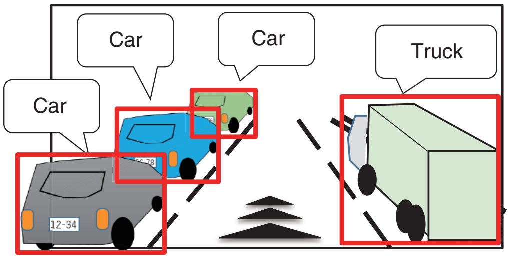


Fig. 3. Detection of vehicles from a frame image (car, truck, or bus).

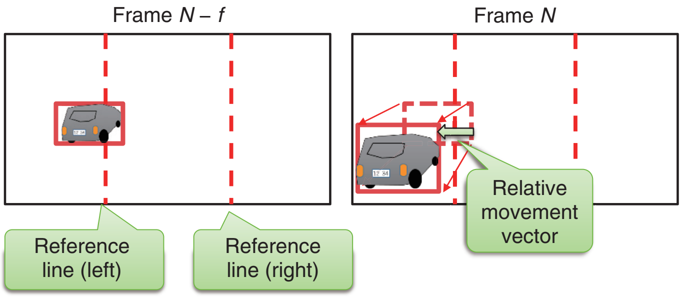


Fig. 4. Counting overtaken vehicles using a reference line.

by tracking is expressed as a relative movement vector. When the relative movement vector (of one of the other vehicles) from the road-perspective vanishing point to the outside of the frame intersects a reference line, it is determined that the observation vehicle has overtaken this other vehicle. In this manner, overtaken vehicles are counted.

(4) Calculating the travel distance

The distance traveled by the observation vehicle is calculated from the vehicle-speed information.

(5) Detecting a lane-specific traffic jam

Unfortunately, the threshold for the number of overtaken vehicles per evaluation timeframe is not clear. Therefore, we instead used the density of the overtaken vehicles (vehicle density) as a temporary solution because it is reasonable to think that the distance between vehicles is shorter in a traffic jam than in normally flowing traffic.

The threshold for the distance occupied by the vehicles in the adjacent lane (called the occupied distance threshold, L_θ) due to a traffic jam is calculated from the number of overtaken vehicles, the length, l_i, of an overtaken vehicle i, and the distance, τ_i, between vehicle i and the vehicle immediately in front of vehicle i. The occupied distance threshold is obtained as

$$L_\theta = \sum_i (l_i + \tau_i). \tag{2}$$

Although vehicle length varies from vehicle to vehicle, we used the following vehicle lengths in our learning model for simplicity. The length of a car $l_{car} = 4.8$ m, that of a truck $l_{truck} = 12.0$ m, and that of a bus $l_{bus} = 12.0$ m. The distance between two vehicles, τ, is set to 5.0 m regardless of the vehicle type because at such a distance it is reasonable to think that the vehicles are packed together but that each vehicle is far enough from the vehicle in front of it to be able to avoid a collision by braking suddenly. Note that the upper limit of traffic speed for the definition of a traffic jam is 10 km/h (2.778 m/s). At this speed, there is a sufficient stopping distance (brake reaction distance + braking distance) between vehicles. The upper limit of the traffic speed for the definition of *minor congestion* is 20 km/h. At this speed, there is not a sufficient stopping distance between vehicles. The reaction time was set to 0.75 s, which is considered a typical value, and the coefficient of friction was set to 0.7 assuming a dry road surface.

Brake reaction distance: $2.778 \text{ m/s} \times 0.75 \text{ s} = 2.08 \text{ m}$

Braking distance: $(10 \times 10) \div (254 \times 0.7) \fallingdotseq 0.56 \text{ m}$

Let x, y, and z respectively denote the number of cars, number of trucks, and number of buses that were overtaken. Then, the occupied length is

$$L_\theta = xl_{car} + yl_{truck} + zl_{bus} + (x + y + z)\tau. \tag{3}$$

If this occupied distance is longer than the observation vehicle's travel distance D, i.e., if

$$L_\theta \geq D, \tag{4}$$

the distance between vehicles is shorter than the defined threshold, it is determined to be a lane-specific traffic jam.

2.4 Other new features

To improve detection accuracy, it is important to prevent object-detection failures caused by backlighting or reflections. For this purpose, a Kalman filter is used to predict where the other vehicles are heading. To determine whether other vehicles are running in the same direction as the observation vehicle or running in the opposite lane, our technology examines whether the video shows the front or back of a vehicle. If there are many vehicles facing towards the camera, the system determines that the other vehicles are in the opposite lane and reverses the front and last vehicles.

3. Field trials

3.1 Overview of the evaluation

The accuracy of our lane-specific traffic-jam-detection technology was evaluated using a test vehicle—running in the Odaiba district, Minato Ward, Tokyo—as the observation vehicle. The vehicle ran on the two routes indicated with the solid red line and dotted blue line in **Fig. 5**. Preliminary investigations indicated that there is a high probability of naturally occurring lane-specific traffic jams on these routes. A camera was installed in the upper front of the observation vehicle. The resolution of the video from this camera was 1920 × 1080 pixels. The video was taken at 10 fps and saved. The location information expressed in latitude and longitude was obtained at one-second intervals using GPS (Global Positioning System). The vehicle's speed was determined on the basis of the change in the vehicle's location over time.

3.2 Evaluation of accuracy in detecting lane-specific traffic jams

This section describes the accuracy evaluation conducted using 48 videos (each 60 s long) for situations in which the presence of a lane-specific traffic jam was visually confirmed.

(1) Visual check

We decided that our technology must be able to detect any lane-specific traffic jam with a length of 42 m or longer. We manually extracted video scenes in which the observation vehicle continuously overtook a series of vehicles, such as three cars and one truck. The data in these visually verified scenes were treated as the correct data.

(2) Accuracy evaluation

Table 1 shows a confusion matrix comparing the results of the visual check and the application output. Precision (true positive (TP)/(TP+ false positive (FP))) was 89.7%. The main factor for an FP was the effect of vehicles in parking lots or other premises. This applied to three cases. The main factor for a false negative (FN) was object-detection errors caused when some vehicles in the opposite lane were hidden behind guardrails or other objects, even though a lane-specific traffic jam was present in that lane. This applied to four cases. Factors common to both FP and FN included object-detection errors, errors caused

Fig. 5. Travel routes.

Table 1. Evaluation of accuracy in detecting lane-specific traffic jams.

N = 98

		Application output	
		Lane-specific traffic jam	No lane-specific traffic jam
Visual check	Lane-specific traffic jam	TP: 70	FN: 8
	No lane-specific traffic jam	FP: 8	TN: 12

because we fixed the occupied distance of other vehicles, and inaccurate travel distance of the observation vehicle due to errors in positional information.

3.3 Evaluation of accuracy in estimating the front vehicle in a queue

We conducted an additional evaluation for 70 TP scenes, scenes for which both the visual check and application output indicated the presence of a lane-specific traffic jam. We compared the frame in which the visual check determined that the observation vehicle overtook the front vehicle with the frame in which the application did the same. The comparison results are shown in **Fig. 6**. The average difference in time between the two frames was 0.57 s, the median was 0 s, and the maximum was 4.1 s. The speed limit

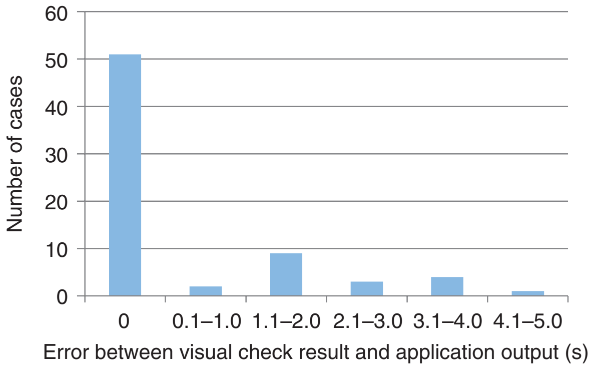


Fig. 6. Accuracy in estimating the front vehicle.

on these routes is 50 km/h. With the distance between cars taken into account, the time required to overtake a car is approximately 0.7 s. Therefore, the average difference of 0.57 s is acceptable. We examined 14 cases in which the error was 1.6 s or more and identified two main factors that caused the large error. The first was that there were errors in the positional information and, as a result, the observation vehicle's travel distance was not calculated accurately (the distance calculation process was either continued or interrupted). This applied to nine cases. The second factor was that cars were incorrectly recognized as trucks or vice versa. This applied to three cases.

4. Conclusion

We developed a lane-specific traffic-jam-detection technology. It uses video data, vehicle speed, and location information that can be collected from a single connected vehicle to determine if there is a traffic jam in each lane and, if there is, identify the front and the last vehicles in the traffic queue. We conducted field trials, in which the accuracy of traffic-jam detection and the estimation of the vehicle at the front of a queue were evaluated on the basis of the video from a front-facing onboard camera and positional information. We plan to improve the accuracy of our technology and study methods for identifying the causes of lane-specific traffic jams.

References

[1] H. Koyama and Y. Shibata, "Dynamic Map Maintenance in Automated Driving," Journal of the Institute of Systems, Control and Information Engineers, Vol. 60, No. 11, pp. 463–468, 2016 (in Japanese).
[2] Website of Central Nippon Highway Engineering Tokyo Company Limited, "Millimeter Wave Radar Type Traffic Monitoring System," https://www.c-nexco-het.jp/tech/english/detail/0048.php
[3] Website of Hitachi Industry & Control Solutions, Ltd., "Hitachi Image Traffic Counter," https://info.hitachi-ics.co.jp/eng/products/gazoutraffics/index.html
[4] D. Watanabe, F. Kanazawa, and J. Sawa, "Development of a Probe Information Collection System Using Smartphones," Proc. of the Technical Study Group, the Ministry of Land, Infrastructure, Transport and Tourism, pp. 221–226, 2013 (in Japanese).
[5] K. Mori, Y. Yokohata, A. Hayashi, T. Hata, and K. Obana, "Traffic Jam Detection in Each Lane and Its Evaluation in Field Trials," Journal of the Information Processing Society of Japan, Vol. 63, No. 1, pp. 163–171, 2022 (in Japanese).
[6] Website of Japan Road Traffic Information Center, FAQ (in Japanese), https://www.jartic.or.jp/s/attention/inquiry/#map 8
[7] Darknet on GitHub, https://github.com/pjreddie/darknet

Kouhei Mori

Research Engineer, NTT Smart Data Science Center.

He received an M.E. from University of Yamanashi in 2009 and joined NTT the same year. He studied ubiquitous services and edge computing at NTT Network Innovation Laboratories. From 2015 to 2018, he was involved in the development of the e-Learning service for corporations and schools at NTT DOCOMO. Since 2018, he has been active in the analysis of data collected from vehicles for driving support at NTT Smart Data Science Center.

Yuki Yokohata

Senior Research Engineer, NTT Smart Data Science Center.

She received an M.S. in earth and planetary science from Nagoya University in 2004 and joined NTT the same year. She studied ubiquitous services at NTT Network Service Systems Laboratories. From 2008 to 2015, she was involved in the development of public wireless local area network services at NTT Communications. Since 2016, she has been active in the development of smart city applications and analysis of data collected from vehicles for driving support at NTT Smart Data Science Center.

Aki Hayashi

Research Engineer, NTT Smart Data Science Center.

She received a B.S. and M.S. in information science from Ochanomizu University, Tokyo, in 2010 and 2012 and joined NTT in 2012. She researched information visualization and data mining at NTT Service Evolution Laboratories. She then received a Ph.D. in science from Ochanomizu University in 2015. From 2016 to 2019, she was involved in the development of voice agent services at NTT DOCOMO. Since 2019, she has been active in the development of smart city applications and analysis of data collected from vehicles for driving support at NTT Smart Data Science Center.

Takahiro Hata

Senior Research Engineer, NTT Smart Data Science Center.

He received an M.Inf. from Kyoto University in 2006 and joined NTT the same year. He studied ubiquitous services at NTT Network Innovation Laboratories. From 2010 to 2013, he was involved in the development of public wireless local area network services at NTT WEST. Since 2013, he has been active in the development of smart city applications and analysis of pedestrian flow and driving data of vehicles at NTT Smart Data Science Center.

Masato Kamiya

Senior Research Engineer, Supervisor, NTT Smart Data Science Center.

He received an M.E. from Nagoya University in 1999 and joined NTT the same year. He studied quality of network services at NTT Network Innovation Laboratories. From 2008 to 2011, he was involved in the development of a next-generation network system at NTT Communications. Since 2011, he has been active in the development of smart city applications and analysis of data collected from vehicles and vessels at NTT Smart Data Science Center.

Technology for Calculating Suddenness Index for Aggregated Values

Aki Hayashi, Yuki Yokohata, Takahiro Hata, Kouhei Mori, and Masato Kamiya

Abstract

We developed a technology for calculating the suddenness index for aggregated values to reduce the amount of computation and communication for video processing needed to detect lane-specific traffic jams. This technology aggregates the number of connected vehicles for each mesh on a map and quantifies the degree of deviation from the ordinary state. This article gives an overview of this technology and describes the value it provides and future issues based on the verification conducted for this lane-specific traffic-jam use case in field trials.

Keywords: statistical analysis, CAN, traffic-jam detection

1. Issues regarding detection of lane-specific traffic jams

NTT Smart Data Science Center aims to provide advanced driving support and optimize traffic flow by detecting lane-specific traffic jams using video analysis, thus providing more detailed traffic-jam information faster than conventional services. To identify the exact location and length of a lane-specific traffic jam, dashcam videos of connected vehicles running in the lane adjacent to the congested lane are collected into the cloud where image processing [1] is executed. This process presents a problem in that a large amount of communication is required to collect the video data and a large amount of computation is required to analyze the data. To solve this problem, NTT Smart Data Science Center developed a technology for calculating the degree of suddenness of the increase or decrease in the latest number of vehicles compared with the number of vehicles in the ordinary state, which is quantified and called the suddenness index for aggregation values, or just suddenness index, *SI* [2]. It uses controller area network (CAN) data, which feature a small data size, and identifies roads (or meshes* on a map), the videos of which need to be collected and analyzed with high priority. By collecting and analyzing only high-priority videos, it is possible to reduce the amount of both communication and computation.

2. Which roads have high priority for video analysis?

Compared with traffic jams in areas where traffic volumes are predictable, such as areas with little change in traffic volume or where traffic jams occur periodically, lane-specific traffic jams that occur unexpectedly and suddenly are more likely to trigger new accidents. Therefore, it is important to provide advanced driving assistance in areas that are prone to such sudden lane-specific traffic jams.

Traffic jams in the former category include those that occur on arterial roads and right/left-turn lanes, which become congested during the morning and evening rush hours, and at entries to commercial

---

* Mesh: A specific square area on a map. In the field trials, a 150-m^2 area was used as a mesh.

facilities that are crowded on weekends and holidays. Such traffic jams can be predicted on the basis of historical data, and the video-analysis results of past traffic jams have already been accumulated. Therefore, the analysis of videos of such areas has low priority, especially when remaining server resources are scarce.

Lane-specific traffic jams in the latter category include those caused by an extreme reduction in traffic capacity on the road due to an accident or vehicle breakdown, those caused by the opening of a new commercial facility, and those caused by changes in people's daily habits, such as a sudden rise in demand for drive-through services due to the COVID-19 pandemic. Such traffic jams are difficult to predict from historical data. Therefore, it is a matter of high priority to understand the traffic situation in such cases through video analysis. By giving high priority to lane-specific traffic jams that occur unexpectedly, it is possible to reduce the amount of both communication and computation required.

3. Issues with detecting sudden traffic jams

The issues with detecting sudden traffic jams are summarized below.

First, it is necessary to use, as input data, information that is readily available and can be analyzed without incurring a high cost in communication and computation. The proposed technology collects CAN data from connected vehicles and uses the number of vehicles in each mesh as input data. It is necessary to take note of the facts that the reliability of data on the number of vehicles is low and will remain so until connected vehicles become widespread and that it is necessary to reduce the amount of computation needed for detecting a sudden traffic jam to improve real-time performance.

To be able to identify roads encountering sudden traffic jams, the number of vehicles in the ordinary state (normal traffic without sudden congestion) should be learned from data on the number of vehicles in each mesh in each hour segment and the time series of that data. In doing so, it is important to take into account the facts that the ordinary state varies from road to road and that the number of vehicles changes periodically, for example, depending on the day of the week, hour of the day, and combination of the two.

4. Technology for calculating suddenness index

Our technology can quickly narrow down the spot where a sudden traffic jam is occurring. The input data it uses are the number of vehicles in each mesh, which is collected with Axispot™ [3] based on CAN data. This technology can detect the moment when the number of vehicles suddenly increases in comparison with the number of vehicles in the ordinary state.

The flow for calculating *SI* is shown in **Fig. 1**. To take multiple types of periodicity into account, the technology calculates the ordinary state (hereafter, referred to as statistical information) of each mesh from various data: the average, standard deviation and time-scale reliability of the number of vehicles in each mesh, which are aggregated for the entire period, by the hour of the day, day of the week, and combination of the two (these are collectively called time granularity). Time-scale reliability is also calculated to determine, from among the results for different time granularities, which should be given high priority. This time-scale reliability is designed in such a way that it is low when the density of connected vehicles is low; thus, the number of vehicles on which information is obtained is excessively small relative to the number of lanes. Specifically, it is calculated on the basis of the number of times that the number of vehicles on which information is obtained increases above a defined threshold. When the latest number of vehicles is input, *SI* is calculated. The input data are the latest number of vehicles (by mesh and by time) and statistical information calculated in advance. First, the suddenness index by time granularity (SI_T) is calculated by comparing the latest number of vehicles with the statistical information for each time granularity. The SI_T is calculated using a method often used in outlier tests [4], as shown in Fig. 1, where $C(t, m)$ is the number of vehicles passing through mesh m at time t, $\mu(T, t, m)$ is the mean number of vehicles at t, at m for time granularity T, and $\sigma(T, t, m)$ is the standard deviation of the same. The SI_T is positive if the latest number of vehicles is greater than the mean and is negative if it is smaller. The larger the standard deviation of the population, the smaller SI_T is. The SI_T is high when a large number of vehicles suddenly accumulate in a place where the number of vehicles is normally small and varies little.

Next, the SI_T for each time granularity is multiplied by the time-scale reliability. The weighted sum of the SI_T values for all time granularities is the final output, called the *SI*. If the value is positive, the number of

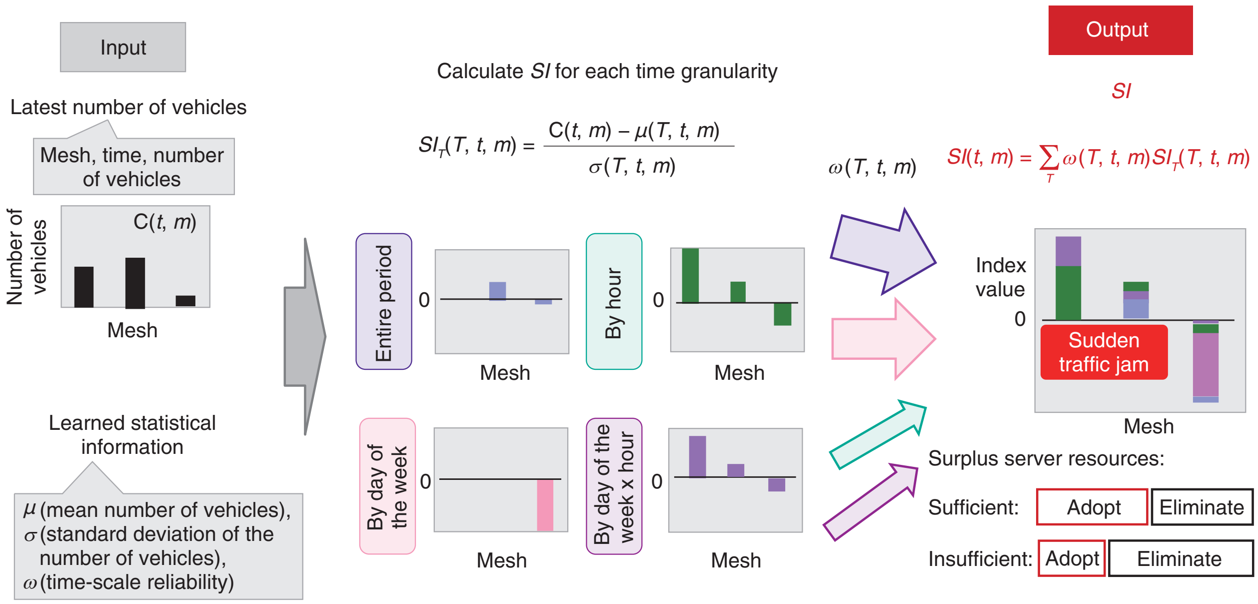


Fig. 1. Calculation method for *SI*.

vehicles is higher than usual at many high-reliability time granularities. This enables flexible selection of the mesh to study. For example, if the server that collects and processes video has sufficient surplus resources, the center and left-most meshes of the chart in the right-hand column in Fig. 1 can be selected. If it does not, only the left-most mesh, which has the highest *SI*, can be selected.

To observe deviations from the ordinary state, it is necessary to know the ordinary state. It has been common to use a complex model, such as Auto Encoder [5], to learn the ordinary state. We adopted a different approach for the lane-specific traffic-jam-detection use case. To minimize the amount of computation and computing time required for prioritizing video-processing-based traffic-jam detection, multiple aggregation time granularities are used, and whether the number of vehicles is greater than in the ordinary state is quantified for each aggregation granularity. A benefit of this approach is that it minimizes the computation time while minimizing the negative effects when the reliability of the data is low.

5. Evaluation of accuracy and performance

Since connected vehicles are still in the early-adoption stage, we believed that a reasonable method for evaluating the accuracy and performance of the proposed technology with real data was to use Global Positioning System (GPS) logs. We calculated *SI* using GPS data collected from 10 taxis every 10 seconds because the occurrence of a traffic jam would reduce vehicle speeds, thus increase the number of GPS logs per unit of time.

A summary of the data used is as follows:

- Mesh size: 110 m^2
- Number of GPS logs: 2,757,003
- Number of meshes: 47,146
- Number of time frames: 9,258
- Total number of all meshes in the aggregation: 1,016,024
- Period: November 27, 2017–January 31, 2018

Figure 2 shows an example of the *SI* calculation for a specific mesh for which the highest *SI* was produced. The *SI* values are shown in red and the number of logs in blue. When the number of logs (in blue) rose suddenly at (1), the *SI* produced with the proposed technology (in red) also increased at (1). This indicates that the technology can be used to detect a sudden traffic jam. When traffic jams occurred repeatedly, as shown with the blue dots at (2) and (3), the *SI* values shown with the red dots at (2) and (3), decreased. This indicates that the use of *SI* gradually eliminates recurring traffic jams of similar magnitude from the target for traffic-jam detection.

Next, we evaluated whether *SI* is a reliable indicator for detecting sudden traffic jams. We visually detected 121 lane-specific traffic jams in videos taken by

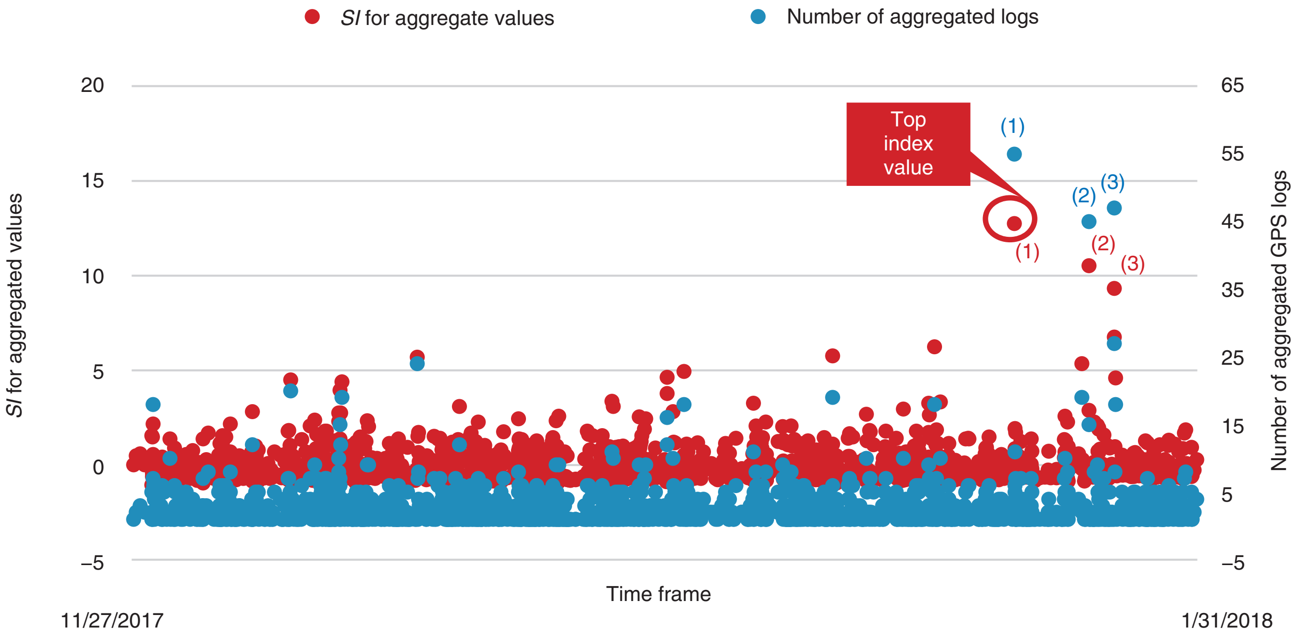


Fig. 2. Example of calculating *SI* for aggregate values.

Table 1. Identification rate of sudden traffic jams.

Effect of aggregation on cost saving (Odaiba)	Effect of aggregation on cost saving (Overall)	Rate of identifying meshes with traffic jams (Odaiba)	Rate of identifying sudden traffic jams (Odaiba)	Rate of identifying chronic traffic jams (Odaiba)
82.2% reduction (1,187/1,444)	72.6% reduction (344,019/473,723)	69.2% identified (9/13)	100% identified (7/7)	42.9% identified (3/7)

taxis, labeled each of them as either sudden or not, and checked whether *SI* is effective for identifying sudden traffic jams. From the taxi data collected over a period of 66 days, we used the data from the first 33 days as training data. We then checked whether *SI* could identify the 7 sudden traffic jams out of the 14 traffic jams that occurred during the remaining 33 days. The results are listed in **Table 1**. The aggregation time at which *SI* became greater than 0 was considered the moment when a sudden traffic jam occurred. The table also shows that, even if 82.2% of the aggregation data taken around Odaiba (the district where the demonstration experiment was conducted) were removed, sudden traffic jams were still identified with a probability of 100%. In contrast, *SI* identified only 42.9% of chronic traffic jams, which are not targeted traffic jams in this use case.

This technology was implemented on the platform used in the field trials, which is presented in another article in this issue. We evaluated the proposed technology in this setting. We examined what reduction in computation time can be achieved if the number of processed meshes (meshes for which video is processed) is reduced by 80% using *SI*. We examined this effect with different numbers of meshes. The results are shown in **Fig. 3**. As the number of meshes increased, the greater the effect of decreasing the number of processed meshes on reducing computing time. When the number of meshes was 400 or 800, the computing time decreased by 70% when using *SI*. However, the time required to compute *SI* remained as short as 0.5 seconds even with 800 meshes.

6. Future outlook

We developed a technology for calculating the

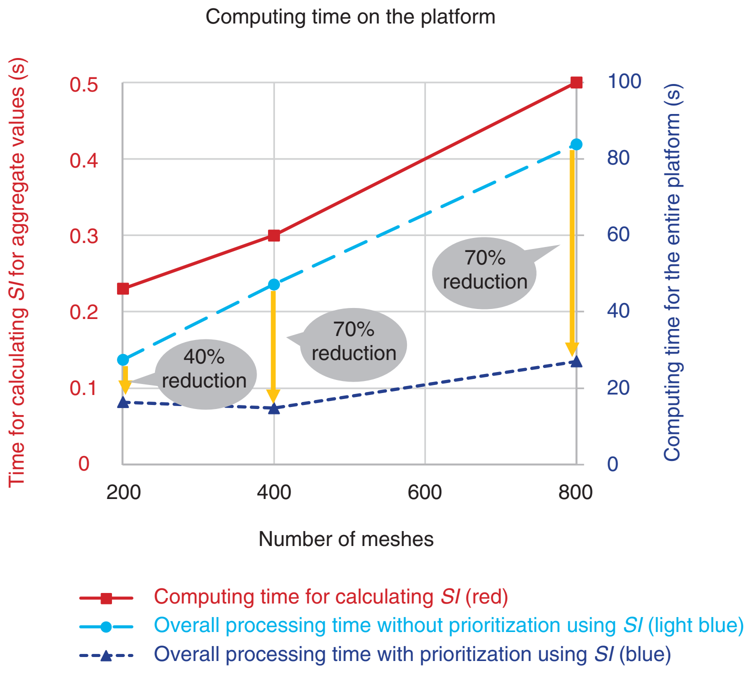


Fig. 3. Effect of using *SI* for aggregate values on reducing computing time on the platform.

suddenness index for aggregated values, which can quickly identify sudden increases in time-series data with a small amount of computation. We plan to use this technology to detect or predict sudden traffic jams on expressways caused by accidents on the basis of information about previous traffic jams on expressways.

References

[1] K. Mori, Y. Yokohata, A. Hayashi, T. Hata, and K. Obana, “Traffic Jam Detection in Each Lane and Its Evaluation in Field Trials,” Journal of the Information Processing Society of Japan, Vol. 63, No. 1, pp. 163–171, 2022 (in Japanese).

[2] A. Hayashi, Y. Yokohata, T. Hata, K. Mori, and K. Obana, “Prioritization of Lane-based Traffic Jam Detection for Automotive Navigation System Utilizing Suddenness Index Calculation Method for Aggregated Values,” 60th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE2021), Tokyo, Japan, pp. 874–880, Sept. 2021.

[3] M. Hanadate, T. Kimura, N. Oki, N. Shigematsu, I. Ueno, I. Naito, T. Kubo, K. Miyahara, and A. Isomura, “Introduction to Axispot™, Real-time Spatio-temporal Data-management System, and Its High-speed Spatio-temporal Data-search Technology,” NTT Technical Review, Vol. 18, No. 1, pp. 22–29, 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202001fa3.html

[4] V. J. Hodge and J. Austin, “A Survey of Outlier Detection Methodologies,” Artificial Intelligence Review, Vol. 22, pp. 85–126, 2004.

[5] S. Nakatsuka, H. Aizawa, and K. Kato, “Defective Products Detection Using Adversarial AutoEncoder,” Proc. of International Workshop on Advanced Image Technology (IWAIT) 2019, Vol. 11049, Singapore, Jan. 2019.

Aki Hayashi

Research Engineer, NTT Smart Data Science Center.

She received a B.S. and M.S. in information science from Ochanomizu University, Tokyo, in 2010 and 2012 and joined NTT in 2012. She researched information visualization and data mining at NTT Service Evolution Laboratories. She then received a Ph.D. in science from Ochanomizu University in 2015. From 2016 to 2019, she was involved in the development of voice agent services at NTT DOCOMO. Since 2019, she has been active in the development of smart city applications and analysis of data collected from vehicles for driving support at NTT Smart Data Science Center.

Yuki Yokohata

Senior Research Engineer, NTT Smart Data Science Center.

She received an M.S. in earth and planetary science from Nagoya University in 2004 and joined NTT the same year. She studied ubiquitous services at NTT Network Service Systems Laboratories. From 2008 to 2015, she was involved in the development of public wireless local area network services at NTT Communications. Since 2016, she has been active in the development of smart city applications and analysis of data collected from vehicles for driving support at NTT Smart Data Science Center.

Takahiro Hata

Senior Research Engineer, NTT Smart Data Science Center.

He received an M.Inf. from Kyoto University in 2006 and joined NTT the same year. He studied ubiquitous services at NTT Network Innovation Laboratories. From 2010 to 2013, he was involved in the development of public wireless local area network services at NTT WEST. Since 2013, he has been active in the development of smart city applications and analysis of pedestrian flow and driving data of vehicles at NTT Smart Data Science Center.

Kouhei Mori

Research Engineer, NTT Smart Data Science Center.

He received an M.E. from University of Yamanashi in 2009 and joined NTT the same year. He studied ubiquitous services and edge computing at NTT Network Innovation Laboratories. From 2015 to 2018, he was involved in the development of the e-Learning service for corporations and schools at NTT DOCOMO. Since 2018, he has been active in the analysis of data collected from vehicles for driving support at NTT Smart Data Science Center.

Masato Kamiya

Senior Research Engineer, Supervisor, NTT Smart Data Science Center.

He received an M.E. from Nagoya University in 1999 and joined NTT the same year. He studied quality of network services at NTT Network Innovation Laboratories. From 2008 to 2011, he was involved in the development of a next-generation network system at NTT Communications. Since 2011, he has been active in the development of smart city applications and analysis of data collected from vehicles and vessels at NTT Smart Data Science Center.

Unsupervised Depth and Bokeh Learning from Natural Images Using Aperture Rendering Generative Adversarial Networks

Takuhiro Kaneko

Abstract

Humans can estimate the depth and bokeh effects from a two-dimensional (2D) image on the basis of their experience and knowledge. However, computers have difficulty in doing this because they logically cannot have such experience and expertise. To overcome this limitation, a novel deep generative model called aperture rendering generative adversarial network (AR-GAN) is discussed. AR-GAN makes it possible to control the bokeh effects on the basis of the predicted depth by incorporating an optical constraint of a camera aperture into a GAN. During training, AR-GAN requires only standard 2D images (such as those on the web) and does not require 3D data such as depth and bokeh information. Therefore, it can alleviate the application boundaries that come from the difficulty in collecting 3D data. This technology is expected to enable the exploration of new possibilities in studies on 3D understanding.

Keywords: generative adversarial networks, unsupervised learning, depth and bokeh

1. Introduction

Humans can estimate the depth and bokeh (shallow depth-of-field (DoF)) effects from a two-dimensional (2D) image on the basis of their experience and knowledge. However, computers have difficulty in doing this because they logically cannot have such experience and expertise. However, considering that, in the future, robots will be able to move around us and the real and virtual worlds will be integrated, it will be necessary to create computers that can act or present information on the basis of 3D data such as depth and bokeh information. Considering that a photo is one of the most frequently used forms of data for recording or saving information, understanding 3D information from 2D images will be valuable for various 3D-based applications to reduce installation cost because it enables using easily available 2D images as input.

Three-dimensional understanding from 2D images has been actively studied in computer vision and machine learning. A successful approach is to learn the 3D predictor using direct or photometric-driven supervision after collecting pairs of 2D and 3D data [1] or sets of multi-view images [2]. This approach demonstrates good prediction accuracy due to the ease of training. However, collecting pairs of 2D and 3D data or sets of multi-view images is not always easy or practical because they require special devices such as a depth sensor or stereo camera.

To reduce the data-collection costs, our team is investigating a fully unsupervised approach for learning 3D representations only from images without any additional supervision. In the study published in the 34th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021) [3], I introduced a new deep generative model called aperture rendering generative adversarial network (AR-GAN), which can learn depth and bokeh effects from standard 2D images such as those on the web. Focus cues that are

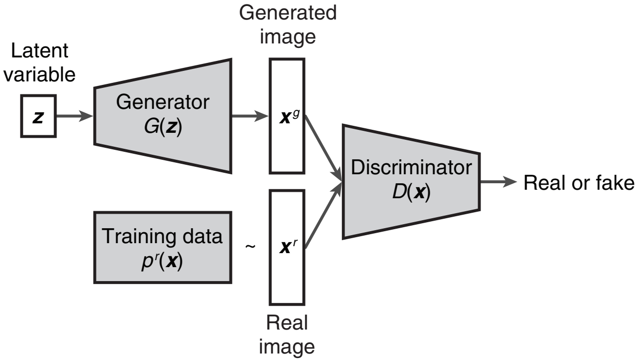


Fig. 1. Architecture of GAN.

inherent in photos but had not been actively studied in previous deep generative models were considered. On the basis of this consideration, our team developed AR-GAN to incorporate aperture rendering (particularly light field aperture rendering [4]) into a GAN [5] (a variant of deep generative models). This configuration allows synthesizing a bokeh image on the basis of the predicted depth and all-in-focus (deep DoF) image using a camera with an optical constraint on the light field.

The rest of this article is organized as follows. In Section 2, I first review two previous studies on which AR-GAN is based: GAN [5] and light field aperture rendering [4]. In Section 3, I explain AR-GAN, which is the main topic of this article. In Section 4, I discuss the experiments on the effectiveness of AR-GAN. In Section 5, I present concluding remarks and areas for future research.

2. Preliminaries

2.1 GAN

GANs [5] can mimic training data without defining their distribution explicitly. This property enables GANs to be applied to various tasks and applications in diverse fields.

As shown in **Fig. 1**, a GAN is composed of two neural networks: a generator $G(z)$ and discriminator $D(\boldsymbol{x})$. These two networks are optimized through a two-player min-max game using an objective function L_{GAN}:

$$L_{\mathrm{GAN}} = \mathbb{E}_{\boldsymbol{x}^r \sim p^r(\boldsymbol{x})}[\log D(\boldsymbol{x}^r)] + \mathbb{E}_{z \sim p(z)}\big[\log\big(1 - D(G(z))\big)\big],$$

where, given a latent variable $z \sim p(z)$, a $G(z)$ attempts to generate an image $\boldsymbol{x}^g = G(z)$ that can deceive a $D(\boldsymbol{x})$ by minimizing L_{GAN}. By contrast, the $D(\boldsymbol{x})$ attempts to distinguish a generated image $\boldsymbol{x}^g$ from a real image $\boldsymbol{x}^r \sim p^r(\boldsymbol{x})$ by maximizing L_{GAN}. Superscripts r and g denote the real and generated data, respectively. Through this adversarial training, a generative distribution $p^g(\boldsymbol{x})$ reaches close to a real distribution $p^r(\boldsymbol{x})$.

2.2 Light field aperture rendering

Light field aperture rendering [4] is a module that simulates an optical phenomenon (particularly bokeh) on a camera aperture in a differentiable manner. Note that such a differentiable property is necessary for deep neural networks (DNNs), such as a $G(z)$ (discussed in Section 2.1), to update the parameters through the backpropagation commonly used for DNNs.

More concretely, as shown in **Fig. 2**, the rendering provides an aperture renderer $R(\boldsymbol{x}_d, \boldsymbol{d})$ that synthesizes a bokeh image $\boldsymbol{x}_s(\boldsymbol{r})$ from an all-in-focus image $\boldsymbol{x}_d(\boldsymbol{r})$ and depth map $\boldsymbol{d}(\boldsymbol{r})$. Here, $\boldsymbol{r}$ indicates the spatial coordinates of the light field on the image plane.

I explain the details in a step-by-step manner. First, a depth map $\boldsymbol{d}(\boldsymbol{r})$ is expanded into a depth map for each view in the light field, i.e., $\boldsymbol{m}(\boldsymbol{r}, \boldsymbol{u})$, using a neural network T:

$$\boldsymbol{m}(\boldsymbol{r}, \boldsymbol{u}) = T(\boldsymbol{d}(\boldsymbol{r})),$$

where $\boldsymbol{u}$ indicates the angular coordinates of the light field on the aperture plane. Subsequently, an all-in-focus image $\boldsymbol{x}_d(\boldsymbol{r})$ is warped into an image for each view of the light field, i.e., $\boldsymbol{l}(\boldsymbol{r}, \boldsymbol{u})$, using the predicted $\boldsymbol{m}(\boldsymbol{r}, \boldsymbol{u})$:

$$\boldsymbol{l}(\boldsymbol{r}, \boldsymbol{u}) = \boldsymbol{x}_d(\boldsymbol{r} + \boldsymbol{u}m(\boldsymbol{r}, \boldsymbol{u})).$$

From this formulation, the left-side images in the light field (5 × 5 images in Fig. 2) represent images

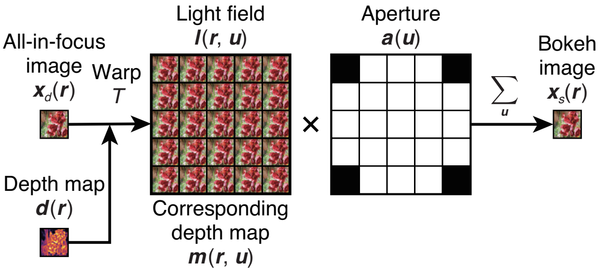


Fig. 2. Processing flow of light field aperture rendering.

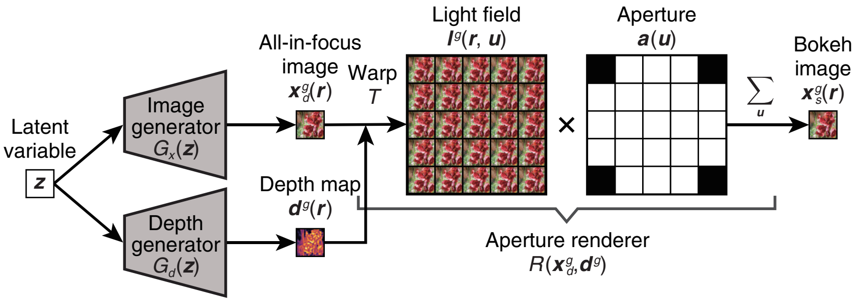


Fig. 3. Processing flow of AR-GAN generator.

when viewing objects from the left side, and the right-side images represent vice versa.

Finally, the $\boldsymbol{l}(\boldsymbol{r}, \boldsymbol{u})$ is integrated using an aperture $\boldsymbol{a}(\boldsymbol{u})$ (an indicator that represents the disk-shaped camera aperture and takes ones for views within the aperture (indicated with white in Fig. 2) and zeroes otherwise (indicated with black in Fig. 2)) to render a bokeh image $\boldsymbol{x}_s(\boldsymbol{r})$:

$$\boldsymbol{x}_s(\boldsymbol{r}) = \textstyle\sum_u \boldsymbol{a}(\boldsymbol{u})\, \boldsymbol{l}(\boldsymbol{r}, \boldsymbol{u}).$$

When an object is on the focal plane, the object's position is consistent regardless of the $\boldsymbol{l}(\boldsymbol{r}, \boldsymbol{u})$. Therefore, no bokeh occurs when the $\boldsymbol{l}(\boldsymbol{r}, \boldsymbol{u})$ is integrated by the above equation. By contrast, when an object is separate from the focal plane, the object's position varies depending on the $\boldsymbol{l}(\boldsymbol{r}, \boldsymbol{u})$. Thus, bokeh occurs in this case. Hereafter, $\boldsymbol{r}$ and $\boldsymbol{u}$ are omitted for simplicity except in necessary cases.

3. AR-GAN

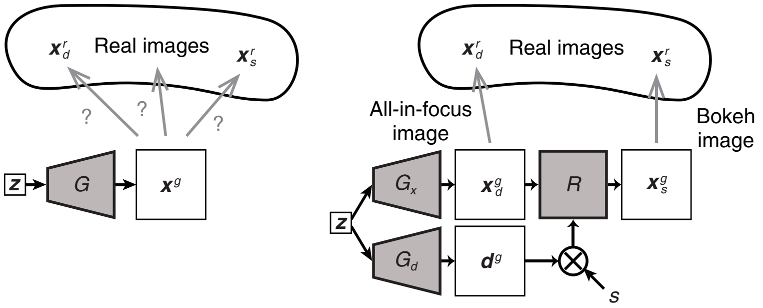
3.1 Problem statement

The problem statement is clarified before explaining the details of AR-GAN. As described in Section 1, AR-GAN is used to learn depth and bokeh effects only from images without additional supervision. In this setting, it is not easy to construct a conditional generator that directly predicts the depth or bokeh effects from an image due to the absence of pairs of 2D and 3D data or sets of multi-view images. Therefore, as an alternative, the aim is to learn an unconditional generator that can generate a tuple of an all-in-focus image $\boldsymbol{x}_d^g$, depth map $\boldsymbol{d}^g$, and bokeh image $\boldsymbol{x}_s^g$ from a latent variable z.

AR-GAN uses focus cues as a clue for addressing this challenge. When the training images are highly biased in terms of bokeh effects (e.g., all training images are all-in-focus), it is difficult to gain focus cues from the images. Therefore, it is assumed with AR-GAN that the training dataset includes various bokeh images. Note that this assumption does not mean that the training dataset contains sets of different bokeh images for each instance. Under this assumption, AR-GAN learns the generator in a wisdom of crowds approach.

3.2 Model architecture

The processing flow of the AR-GAN generator is presented in **Fig. 3**. Given a latent variable z, the AR-GAN generator generates an all-in-focus image



(a) Standard GAN learning

(b) DoF mixture learning

Fig. 4. Comparison between standard GAN learning and DoF mixture learning.

$\boldsymbol{x}_d^g = G_{\boldsymbol{x}}(z)$ and depth map $\boldsymbol{d}^g = G_{\boldsymbol{d}}(z)$ using an all-in-focus image generator $G_{\boldsymbol{x}}(z)$ and a depth generator $G_{\boldsymbol{d}}(z)$, respectively. Subsequently, the aperture renderer $R(\boldsymbol{x}_d, \boldsymbol{d})$ (explained in Section 2.2) synthesizes a bokeh image $\boldsymbol{x}_s^g = R(\boldsymbol{x}_d^g, \boldsymbol{d}^g)$. Using this configuration, AR-GAN makes it possible to generate a tuple of an all-in-focus image $\boldsymbol{x}_d^g$, depth map $\boldsymbol{d}^g$, and bokeh image $\boldsymbol{x}_s^g$ using a camera with an optical constraint on the light field.

3.3 Training method

As shown in Fig. 1, a typical GAN applies a $D(\boldsymbol{x})$ to the final output of the $G(z)$ (i.e., $\boldsymbol{x}_s^g$ in the case of the AR-GAN generator). However, in the AR-GAN generator, three modules, i.e., $G_{\boldsymbol{x}}(z)$, $G_{\boldsymbol{d}}(z)$, and $R(\boldsymbol{x}_d^g, \boldsymbol{d}^g)$ are trainable. Therefore, they compete for roles if there is no constraint. For example, they can fall into an extreme solution (e.g., $R(\boldsymbol{x}_d^g, \boldsymbol{d}^g)$ learns strong bokeh effects and $G_{\boldsymbol{x}}(z)$ learns over-blurred images).

To alleviate this problem, AR-GAN is trained using DoF mixture learning. **Figure 4** illustrates the comparison between the standard GAN learning and DoF mixture learning. In the standard GAN learning shown in Fig. 4(a), the $G(z)$ attempts to cover the overall real image distribution using generated images without any constraint. Consequently, it cannot determine to make a generated image $\boldsymbol{x}^g$ close to a real all-in-focus image $\boldsymbol{x}_d^r$ or a real bokeh image $\boldsymbol{x}_s^r$ (indicated with question marks "?" in Fig. 4(a)).

By contrast, as shown in Fig. 4(b), in DoF mixture learning, the AR-GAN generator attempts to represent the real image distribution using generated images, the bokeh degrees of which are adjusted by a scale factor s. More concretely, the GAN objective function presented in Section 2.1 is rewritten as follows:

$$L_{\text{AR-GAN}} = \mathbb{E}_{\boldsymbol{x}^r \sim p^r(\boldsymbol{x})}[\log D(\boldsymbol{x}^r)] + \mathbb{E}_{z \sim p(z), s \sim p(s)} \left[\log\left(1 - D\left(R\left(G_{\boldsymbol{x}}(z), sG_{\boldsymbol{d}}(z)\right)\right)\right)\right],$$

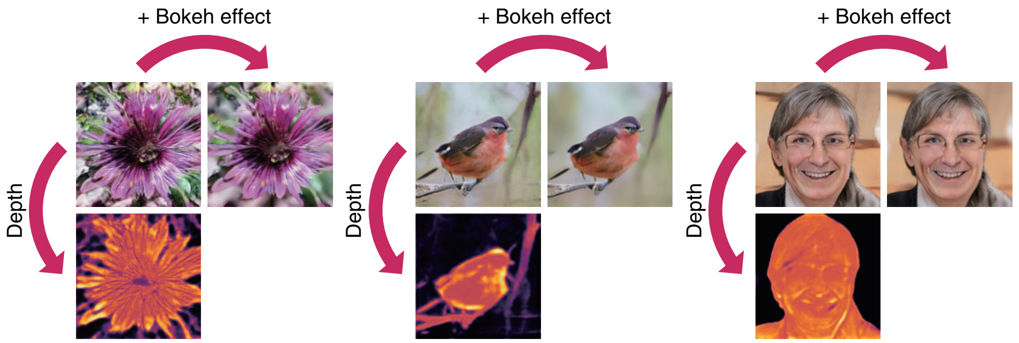
where $s \in [0, 1]$; when $s = 0$, an all-in-focus image $\boldsymbol{x}_d^g$ is generated, whereas when $s = 1$, a bokeh image $\boldsymbol{x}_s^g$ is rendered. Intuitively, the aperture renderer $R(\boldsymbol{x}_d^g, \boldsymbol{d}^g)$, which has an optical constraint on the light field, functions as a bokeh image prior. This prior encourages a generated all-in-focus image $\boldsymbol{x}_d^g$ to approximate a real all-in-focus image $\boldsymbol{x}_d^r$ (indicated by the "All-in-focus image" in Fig. 4(b)) and promotes a generated bokeh image $\boldsymbol{x}_s^g$ to mimic a real bokeh image $\boldsymbol{x}_s^r$ (indicated by the "Bokeh image" in Fig. 4(b)). Consequently, $\boldsymbol{d}^g$, which connects $\boldsymbol{x}_d^g$ and $\boldsymbol{x}_s^g$, is also optimized. In this manner, the DoF mixture learning allows optimizing $G_{\boldsymbol{x}}(z)$, $G_{\boldsymbol{d}}(z)$, and $R(\boldsymbol{x}_d^g, \boldsymbol{d}^g)$ together under an optical constraint.

A remaining challenge specific to unsupervised depth and bokeh learning is the difficulty in distinguishing whether blur occurs ahead of or behind the focal plane. For this challenge, on the basis of the observation that the focused image tends to be placed at the center of a photo, AR-GAN uses the center focus prior, which encourages the center to be focused while promoting the surroundings to be behind the focal plane. In practice, this prior is only used at the beginning of training to determine the learning direction.

4. Experiments

4.1 Image and depth synthesis

The previous AR-GAN study [3] demonstrated the utility of AR-GAN using various natural image datasets, including flower (Oxford Flowers [6]), bird (CUB-200-2011 [7]), and face (FFHQ [8]) datasets.



(a) Trained with flower images (b) Trained with bird images (c) Trained with face images

Fig. 5. Examples of generated images and depth maps.

(a) Input (all-in-focus image) (b) Output 1 (bokeh image) (c) Output 2 (depth map)

Fig. 6. Examples of bokeh rendering and depth prediction.

The implementation details are omitted because of space limitations. See that AR-GAN study [3] if interested in the implementation details.

Figure 5 shows examples of generated images and depth maps. AR-GAN succeeds in generating a tuple of an all-in-focus image (upper left), bokeh image (upper right), and depth map (lower left) in every setting. For example, in Fig. 5(a), the background is blurred while the foreground is unchanged in bokeh conversion (the conversion from the upper left to upper right). In depth prediction (the transformation from the upper left to lower left), the depth map (lower left) corresponding to the image (upper left) is successfully predicted. A light color indicates the foreground while a dark color indicates the background. Recall that the training data are only 2D images, and depth and bokeh effects are not provided as supervision. In this manner, learning depth and bokeh effects only from 2D images is the main strength of AR-GAN.

4.2 Application to bokeh rendering and depth prediction

As discussed in Section 3.1, AR-GAN learns an unconditional generator that generates a tuple of an all-in-focus image $\boldsymbol{x}_d^g$, depth map $\boldsymbol{d}^g$, and bokeh image $\boldsymbol{x}_s^g$ from a latent variable $\boldsymbol{z}$. Therefore, it cannot be directly used to convert a given image to the bokeh image or depth. However, AR-GAN can generate sets of all-in-focus and bokeh images or sets of all-in-focus images and depth maps artificially and abundantly by randomly changing the latent variable. By using these data, we can learn a bokeh renderer (i.e., a converter that converts an all-in-focus image to a bokeh image) and depth predictor (i.e., a predictor that predicts a depth map from an image) in a supervised manner.

Figure 6 shows example results obtained with the bokeh renderer and depth predictor mentioned above. A photo I took was used as an input (Fig. 6(a)). The bokeh renderer synthesizes a bokeh image (Fig. 6(b)), and the depth predictor predicts a depth map from the input image (Fig. 6(c)). Similar to the results in Fig. 5, the background is blurred while the foreground remains unchanged in the bokeh conversion (the conversion from (a) to (b)), and the depth map corresponding to the input image is predicted in the depth prediction (the transformation from (a) to (c)).

Note that the data required for training the bokeh renderer and depth predictor are only the data generated by AR-GAN, and no additional data are needed. That is to say, in this setting, we can learn a bokeh renderer and depth predictor in a fully unsupervised manner, similar to AR-GAN. This is a strength of an AR-GAN-based approach.

5. Conclusion and future work

This article explained AR-GAN, which is a new deep generative model enabling the unsupervised learning of depth and bokeh effects only from natural images. Since we live in the 3D world, human-oriented computers are expected to understand the 3D world. For this challenge, AR-GAN is effective because it can eliminate the requirement of 3D data during training. AR-GAN is expected to enable the exploration of new possibilities in studies on 3D understanding.

AR-GAN will also be useful for many applications in various fields such as environmental understanding in robotics, content creation in advertisements, and photo editing in entertainment. For example, AR-GAN can learn a data-driven model from collected images. Using this strength, a data-driven bokeh renderer reflecting a famous photographer can be constructed if we can collect his/her photos. Thus, AR-GAN can be used to obtain more natural and impactful bokeh images and enrich the functionality of photo-editing applications (e.g., smartphone applications for social media).

Future work includes further improvement of depth and bokeh accuracy since unsupervised learning of depth and bokeh effects is an ill-posed problem, and there is room for improvement. Our team is tackling this challenge, and my latest paper [9] has been accepted to CVPR 2022. Due to space limitations, details of this are omitted. Please check my latest paper [9] if interested in the details.

References

[1] D. Eigen, C. Puhrsch, and R. Fergus, “Depth Map Prediction from a Single Image Using a Multi-scale Deep Network,” Proc. of the 27th International Conference on Neural Information Processing Systems (NIPS 2014), pp. 2366–2374, Montreal, Canada, Dec. 2014.

[2] R. Garg, V. Kumar B. G., G. Carneiro, and I. Reid, “Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue,” Proc. of the 14th European Conference on Computer Vision (ECCV 2016), pp. 740–756, Amsterdam, The Netherlands, Oct. 2016.

[3] T. Kaneko, “Unsupervised Learning of Depth and Depth-of-field Effect from Natural Images with Aperture Rendering Generative Adversarial Networks,” Proc. of the 34th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021), pp. 15679–15688, Virtual, June 2021.

[4] P. P. Srinivasan, R. Garg, N. Wadhwa, R. Ng, and J. T. Barron, “Aperture Supervision for Monocular Depth Estimation,” Proc. of CVPR 2018, pp. 6393–6401, Salt Lake City, USA, June 2018.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” Proc. of NIPS 2014, pp. 2672–2680, Montreal, Canada, Dec. 2014.

[6] M. Nilsback and A. Zisserman, “Automated Flower Classification over a Large Number of Classes,” Proc. of the Sixth Indian Conference on Computer Vision, Graphics & Image Processing (ICVGIP 2008), pp. 722–729, Bhubaneswar, India, Dec. 2008.

[7] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, “The Caltech-UCSD Birds-200-2011 Dataset,” Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.

[8] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” Proc. of CVPR 2019, pp. 4401–4410, Long Beach, USA, June 2019.

[9] T. Kaneko, “AR-NeRF: Unsupervised Learning of Depth and Defocus Effects from Natural Images with Aperture Rendering Neural Radiance Fields,” CVPR 2022, pp. 18387–18397, New Orleans, USA, June 2022.

Takuhiro Kaneko

Distinguished Researcher, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.E., M.S., and Ph.D. from the University of Tokyo in 2012, 2014, and 2020. He joined NTT Communication Science Laboratories in 2014 and has been a distinguished researcher since 2020. His research interests include computer vision, signal processing, and machine learning. He is currently working on image synthesis, speech synthesis, and voice conversion using deep generative models such as GANs. He received the Hatakeyama Award from the Japan Society of Mechanical Engineers, the ICPR 2012 Best Student Paper Award in the 21st International Conference on Pattern Recognition (ICPR 2012), and the Dean's Award for Best Doctoral Thesis from the University of Tokyo. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan, the Information Processing Society of Japan (IPSJ), and the Acoustic Society of Japan (ASJ).

Report of the 9th ITU-T TSAG (Telecommunication Standardization Advisory Group) Meeting

Noriyuki Araki

Abstract

The 9th meeting of the Telecommunication Standardization Advisory Group (TSAG) of the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T), the final meeting of the study period (2017–2021) for the World Telecommunication Standardization Assembly (WTSA-20), was held in an online conference from January 10 to 17, 2022. The 4th Inter-regional Meeting, a preparatory meeting for WTSA-20, was also held on January 6 before the TSAG meeting. This article describes the main results of the 9th TSAG meeting.

Keywords: ITU, TSAG, standardization

1. Introduction

From January 10 to 17, 2022, the 9th meeting of the Telecommunication Standardization Advisory Group (TSAG) of the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) was held remotely with 206 participants from 49 countries. From Japan, the Ministry of Internal Affairs and Communications served as the Head of Delegation for Japan, with 15 representatives from Japanese companies and organizations (National Institute of Standards and Technology (NICT), NTT, KDDI, NEC, The ITU Association of Japan, Hitachi, Mitsubishi, and the Telecommunication Technology Committee (TTC)).

2. Outline of the TSAG meeting

The 9th TSAG meeting consisted of plenary sessions for overall deliberation and Rapporteur Group (RG) meetings for detailed examination of each issue, with two online sessions per day. As this was the final TSAG meeting of the study period (2017–2021) for the World Telecommunication Standardization Assembly (WTSA-20), the closing plenary was allocated for two days on January 14 and 17.

The plenary sessions were led by TSAG Chair Bruce Gracie (Ericsson, Canada), and the RG meetings were led by four Rapporteurs: RG on Work Programme (RG-WP), RG on Working Methods (RG-WM), RG on Strengthening Collaboration (RG-SC), and RG on Review of Resolutions (RG-RR). RG on Standardization Strategy (RG-StdsStrat) had completed the discussion of major issues, and there was no discussion at this meeting. From Japan, Ms. Miho Naganuma (NEC) participated as an RG-WP Rapporteur.

The online conference was held at 13:00–16:00 Geneva time (21:00–24:00 Japan time), with 1 session of 90 minutes and 2 sessions a day as the core time, taking into account the time difference between participants.

We used Zoom for this meeting as a remote meeting tool. The ITU has its own remote meeting tool, MyMeetings, which is built on open source, but for large meetings with more than 200 participants, Zoom will be used for the time being. From the viewpoint of improving accessibility, captioning in English was provided at all meetings and simultaneous interpretation in six official languages of the United Nations was provided at plenary sessions in core time.

3. Inter-regional Meeting for WTSA

On January 6, the week before the 9th TSAG meeting, the Inter-regional Meeting (IRM), a preparatory meeting for WTSA, was held by representatives of six regional telecommunications organizations (the Asia-Pacific Telecommunity (APT), League of Arab States/Arab Standardization Team, African Telecommunications Union (ATU), European Conference of Postal and Telecommunications Administrations (CEPT), Inter-American Telecommunication Commission (CITEL), and Reginal Commonwealth in the Field of Communications (RCC)), and a summary report was given to the TSAG plenary. Sixty-one proposals for revision or withdraw of existing WTSA resolutions, 12 proposals for new resolutions, and 17 proposals for revision of ITU-T A-Series Recommendations were submitted as their common proposals to WTSA-20 for discussion.

The purpose of the IRM is to exchange information in advance to promote mutual understanding of the proposals, including a comparative analysis of the proposals made by regional organizations for resolutions and recommendations of the WTSA.

From APT, Mr. Yoichi Maeda of the TTC presented a progress report as the chair of the APT WTSA Preparatory Meeting and provided the latest information on the 29 focal points of the APT Co-Proposals for deliberation toward the unification of the draft resolutions at the WTSA.

4. Discussions toward WTSA-20

WTSA-20, the highest decision-making meeting for the ITU-T's operational policies, was scheduled to be held in Hyderabad, India in November 2020, but was postponed twice due to the COVID-19 pandemic. The meeting in India was finally cancelled, and the WTSA was held from March 1 to 9 in 2022 in Geneva, where the ITU headquarters are located. The venues were the Geneva International Conference Center adjacent to the ITU headquarters and the Montbrillant building of the ITU headquarters.

On February 28, just before WTSA-20, the Global Standards Symposium (GSS) was held. The theme of this GSS was "International standards for enabling digital transformation and achieving sustainable development goals," and WTSA-20 attendees were expected to participate in the GSS.

Under the current WTSA rules, votes at the WTSA-20 plenary session can only be taken by physical participants representing each country, and remote participants cannot participate in the vote. However, in consideration of the impact of the COVID-19 pandemic, participants were able to participate in discussions at all WTSA-20 meetings in a remote online format and were able to participate in monitoring of plenary sessions.

5. Meetings of Focus Group, Joint Coordination Activities, and Ad-Hoc Group

The activities of the Focus Group (FG), Joint Coordination Activities (JCA), and Ad-Hoc Group (AHG) were reported. The results of the FG's deliberations will be transferred to the relevant Study Groups (SGs), which will be considered as new issues handled by the ITU-T. This will have an impact on the structure of Questions for each SG in the next study period and will be important to understand the trends of future deliberations.

(1) FG on Quantum Information Technology for Networks (FG-QIT4N)

The final activity report of FG-QIT4N on quantum information and communication technology was presented, and the report was approved. This FG was established in September 2019 under the leadership of TSAG and worked until December 2021, producing nine outcome documents. To promote the specific standardization of the outcome document, FG members proposed to plan a briefing session for each SG to facilitate the transition of the outcome document to the related SG and agreed to send a liaison document to all SGs to that effect.

The outcome documents will be compiled into nine documents, including glossary terms, use cases, the quantum key distribution network protocol, and network technology. They will be transferred to SG13, SG17, SG11, and SG15, which are closely related to each other, for further study. This FG was managed by three co-chairs from Russia, the United States, and China.

(2) JCA on Digital COVID-19 Certificates (JCA-DCC)

Regarding the JCA-DCC whose establishment was agreed at the previous TSAG meeting, it was agreed to issue a liaison document to publicize the establishment of the JCA, the purpose of its activities, and the deadline for applications to participate in the JCA of the end of February. The scope of the JCA is to coordinate the work of standardizing digital COVID-19 certificates among the relevant ITU-T SGs, external organizations, and forums; promote compatible data architectures for data sharing; and promote interoperability, agility, and security for all parties involved with users. The JCA-DCC is chaired by Mr. Heung Youl Youm (Korea, Chairman of SG17), and the first meeting will be held electronically in May 2022.

(3) FG on Testbeds Federations for IMT-2020 and beyond (FG-TBFxG)

The establishment of an FG-TBFxG was reported, with SG11, which handles signaling protocols and test specifications, as its parent. This FG functions as a platform to harmonize testbed specifications between standards developing organizations/Fora, develops an application programming interface (API) along the testbed federation reference model defined in Recommendation ITU-T Q.4068 developed in collaboration with European Telecommunications Standards Institute (ETSI) Technical Committee on Core Network and Interoperability Testing, and defines a set of federated testbeds and API use cases.

6. RG meeting

The following is a summary of the main issues discussed in the RG established at TSAG. The report of each RG is approved at the closing plenary of TSAG and reflected as part of the TSAG meeting report at WTSA.

(1) RG-WP

The RG-WP is a group that deals with issues related to SG restructuring, and was given a two-session time frame for deliberation. This RG reviewed all SG activity reports, sought endorsement of the proposed agenda in plenary, and compiled discussions on SG restructuring toward WTSA-20.

It was assumed that the SG configuration would be maintained at the current configuration with 11 SGs at the WTSA-20. Full-scale reorganization will be discussed in the lead-up to WTSA-24. To accelerate this discussion, the Correspondence Group (CG) was established under the RG-WP to analyze and study the optimal distribution of SG configuration. Mr. Philip Rushton (Department for Digital, Culture, Media & Sport, United Kingdom), Chairman of CG, reported the action plan of CG and agreed on the action plan for analysis of SG restructuring.

The action plan for the SG restructuring analysis was discussed during the meeting with two additional editing sessions. This action plan aims to thoroughly review the potential restructuring options of the ITU-T based on empirical analysis, with a view to approving the SG restructuring plan at WTSA-24. In advancing the action plan, it was agreed that the definition of the key performance indicators (KPIs)/metrics to be collected would be clarified, the priorities of the various KPIs/metrics to be collected and the timing of implementation of KPIs/statistics would be clarified, and consideration would be given to the funds to be considered.

(2) RG-WM

The RG-WM reviews the WTSA Resolution 1 and A-Series Recommendations (Resolution 32, Recommendation A.1, Recommendation A.7, Recommendation A.8, etc.), which stipulate various ITU-T work procedures and rules. There were many issues to discuss at this meeting, and the RG-WM was allocated a three-session time frame. Regarding the proposal to revise Chapter 5.3 of the JCA proposed by Korea in ITU-T Recommendation A.1, “Working Method of ITU Telecommunication Standardization Sector,” the necessity was not fully understood based on the discussions at the RG-SC, and the proposal to revise JCA was not agreed upon.

(3) RG-SC

The RG-SC is studying ways to strengthen cooperate with other standards bodies and measures. At this meeting, we discussed liaison activities to strengthen cooperation between sectors within the ITU and other standardization organizations such as International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) Joint Technical Committee 1 and oneM2M. It was agreed in plenary on the revised draft recommendation of A.5 “General procedure for including references to other organizations’ documents in ITU-T Recommendations” as a WTSA-related issue after 1 drafting session, and it was decided to propose the revised draft recommendation to WTSA-20. It was also agreed on a draft recommendation to revise ITU-T Recommendation A. 23, “Cooperation with the International Organization for Standardization on Information Technology (ISO) and the International Electrotechnical Commission (IEC) - Appendix II: Best Practices” and proposed it to WTSA-20.

7. Schedule of future meetings

The first meeting of TSAG for the new study period (2022–2024) is scheduled to be held in Geneva on December 12–16, 2022.

Noriyuki Araki

Director, Standardization Office, Research and Development Planning Department, NTT.

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 of ITU-T SG6 from 2006 to 2008 and the rapporteur of Question 17 of 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 was the vice-chairman of ITU-T SG15 from 2013 to 2022. He also contributes to the activities of IEC Technical Committee 86 (fiber optic systems). He received the ITU-AJ award from the ITU Association of Japan in 2017. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

External Awards

IPSJ Outstanding Paper Award

Winners: Takashi Koide, NTT Security Japan; Daiki Chiba, NTT Security Japan; Mitsuaki Akiyama, NTT Social Informatics Laboratories; Katsunari Yoshioka, Yokohama National University; Tsutomu Matsumoto, Yokohama National University
Date: March 30, 2022
Organization: Information Processing Society of Japan (IPSJ)

For "Understanding the Fake Removal Information Advertisement Sites."
Published as: T. Koide, D. Chiba, M. Akiyama, K. Yoshioka, and T. Matsumoto, "Understanding the Fake Removal Information Advertisement Sites," Journal of Information Processing, Vol. 29, pp. 392–405, 2021.

IEEE Senior Member

Winner: Takayuki Ogasawara, NTT Basic Research Laboratories
Date: April 30, 2022
Organization: The Institute of Electrical and Electronics Engineers (IEEE)

IEEE Senior Membership is an honor bestowed only to those who have made significant contributions to the profession.

Honorable Mention Award

Winners: Jack Jamieson, NTT Communication Science Laboratories; Daniel A. Epstein, University of California Irvine; Yunan Chen, University of California Irvine; Naomi Yamashita, NTT Communication Science Laboratories
Date: May 5, 2022
Organization: ACM Conference on Human Factors in Computing Systems (CHI) 2022

For "Unpacking Intention and Behavior: Explaining Contact Tracing App Adoption and Hesitancy in the United States."
Published as: J. Jamieson, D. Epstein, Y. Chen, and N. Yamashita, "Unpacking Intention and Behavior: Explaining Contact Tracing App Adoption and Hesitancy in the United States," Proc. of CHI 2022, New Orleans, USA, Apr./May 2022.

Best Paper Award

Winners: Koji Yamamoto, Kyoto University; Takayuki Nishio, Kyoto University; Masahiro Morikura, Kyoto University; Hirantha Abeysekera, NTT Access Network Service Systems Laboratories
Date: May 7, 2022
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE) Communications Society

For "Stochastic Geometry Analysis of Inversely Proportional Carrier Sense Threshold and Transmission Power for WLAN Spatial Reuse."
Published as: K. Yamamoto, T. Nishio, M. Morikura, and H. Abeysekera, "Stochastic Geometry Analysis of Inversely Proportional Carrier Sense Threshold and Transmission Power for WLAN Spatial Reuse," IEICE Trans. Commun., Vol. E104.B, No. 10, pp. 1345–1353, 2021.