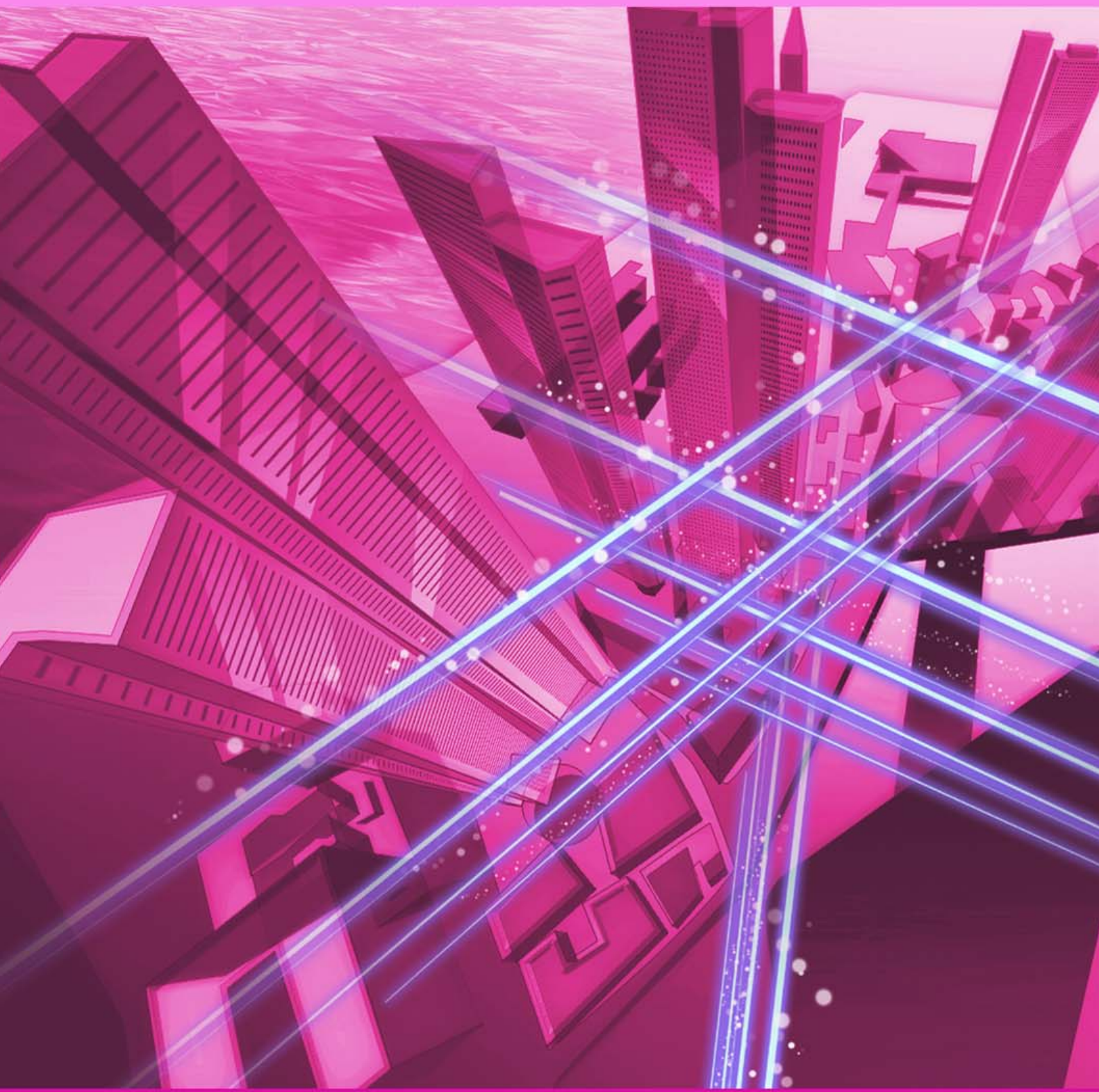
NTT Technical Review

11
2017



November 2017 Vol. 15 No. 11

NTT Technical Review

November 2017 Vol. 15 No. 11

and Digital Operations Strategies

- World’s Largest Transmission Capacity with Standard Diameter Multi-core Optical Fiber—Accelerated Multi-core Fiber Application Using Current Standard Technology

External Awards/Papers Published in Technical Journals and Conference Proceedings

- External Awards/Papers Published in Technical Journals and Conference Proceedings

Making Near-future Predictions about the Flow of People, Things, and Information through Big Data Analysis—Gaining an Edge by the Ability to Produce Value

Naonori Ueda
NTT Fellow, Head of Ueda Research Laboratory and Director of Machine Learning and Data Science Center, NTT Communication Science Laboratories

Overview

The 2016 White Paper on Information and Communications in Japan issued by the Ministry of Internal Affairs and Communications states that the proactive use of information and communication technology such as the Internet of Things and artificial intelligence has the potential to accelerate economic growth in Japan, and the key to this growth will be the collection and use of big data. Against this background, there are high expectations for research achievements in machine learning. NTT Fellow Naonori Ueda of NTT Communication Science Laboratories has announced a string of Japan-first and world-first achievements in machine learning analysis and technology. We asked him about important aspects of research and his frame of mind as a researcher.

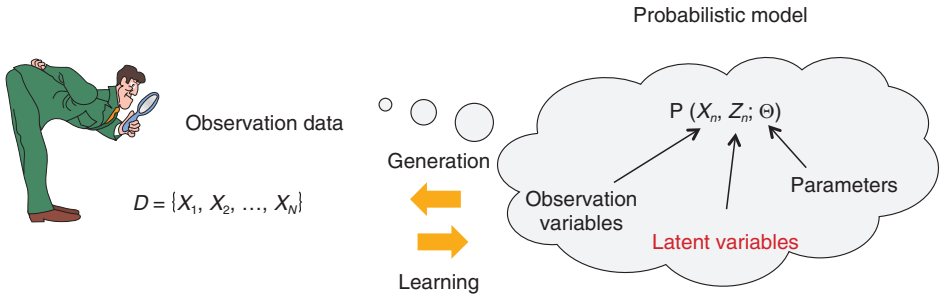
Keywords: machine learning, big data, spatio-temporal data analysis

How an open recruitment for researchers—and criticism too—led to impressive achievements

—Dr. Ueda, please tell us about your current research activities.

In brief, I research machine learning technology. When I first started on this path more than 20 years ago, I was involved in core theoretical research focused on machine learning in the field of artificial intelligence (AI), which has recently taken on a broader meaning. Over time, I expanded my research into applications and practical uses of machine learning in conjunction with big data and the Internet of Things (IoT).

Machine learning technology consists of various types of learning methods. One such method is pattern recognition as used in character recognition. This



Statistical machine learning assumes a probabilistic model including latent variables lying behind observation data and describing the data generation process. It learns the model from the observation data and learns (estimates) the model's parameters and latent variables. Finally, it extracts hidden information from the observation data via those latent variables.

Fig. 1. Statistical machine learning technology.

type of learning involves, for example, teaching a computer how to identify a cat by declaring "This is a cat" when presenting a photograph of a cat. Another type of learning is *clustering* as a basic process of data analysis, which involves judging on one's own what is common and different between a cat and some other entity.

However, in keeping with today's information society, we can say that machine learning is more than just pattern recognition; it is a technology for extracting latent information behind data. Statistical learning, which is my area of specialization, treats the data generation process as a probabilistic model. It learns the model from observation data and extracts latent information via this model (**Fig. 1**).

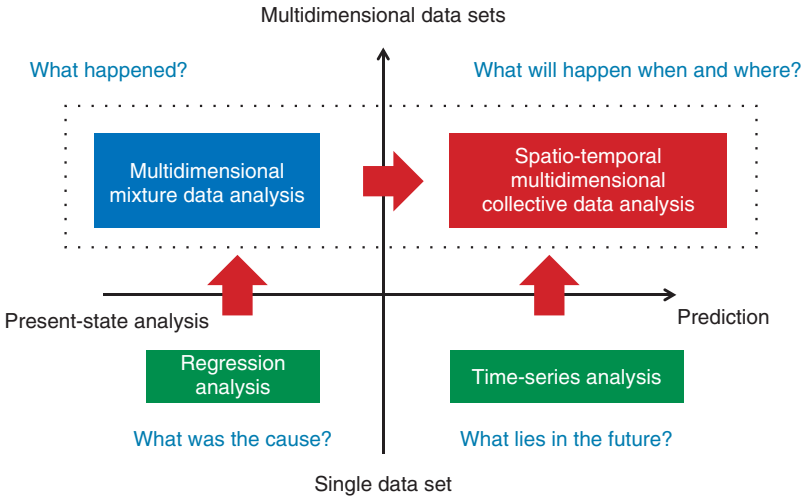
Presently, as director of the Machine Learning and Data Science Center, I am promoting research and development (R&D) that will enable machine learning technology within the NTT AI technology brand corevo® to be put to good use in the real world. Specifically, this is R&D on a system for smoothing out the flows of people and traffic by predicting congestion arising from unforeseen circumstances using a technology called ambient intelligence and providing that information beforehand in real time. This system could be used to provide high-speed evacuation guidance in the event of a disaster.

The technologies used here, and in particular, crowd navigation technology that can respond to unforeseen circumstances, constitute a challenging research theme since they have yet to be used in the real world. We have named this kind of analysis technology targeting such spatio-temporal data typified by people and traffic flows *spatio-temporal multidimensional collective data analysis* (**Fig. 2**) [1].

With conventional technology, it has been difficult to make near-future predictions that specify time and space such as when and where some type of event such as congestion will occur. In contrast, the spatio-temporal multidimensional collective data analysis that we are developing can make a near-future prediction of the period (time) and place (space) of an event that will take place several hours into the future from big data generated sometime in the past. Additionally, as the IoT era gathers momentum, our aim is to optimize the flow of information gathered from a massive number of diverse types of sensors through repeated observation, analysis, prediction, and provision (guidance) in real time. This optimization process should prove useful in preventing congestion itself, controlling communication traffic at the time of a sports or music event, controlling traffic according to the degree of congestion, and guiding visitors in shopping malls, airports, and other places.

—What led you to take on this research theme?

The catalyst was an open recruitment for researchers within the NTT laboratories more than 20 years ago. At that time, in the middle of the AI boom, I applied for a position in AI research at NTT Communication Science Laboratories. To give some background here, it had been decided to set up a basic research laboratory at Keihanna Science City in



Data analysis can be organized in terms of an axis representing present-state analysis or prediction and an axis representing a single set of data or multiple heterogeneous (multidimensional) data sets. Conventional data analysis mostly consists of present-state analysis and prediction for a single set of data. The IoT era, however, features the ongoing provision of information from a wide variety of sensors. Spatio-temporal multidimensional collective data analysis technology can create new value from multidimensional spatio-temporal data linked to time and space. The aim here is to create new value by inferring what will happen when and where from multidimensional big data.

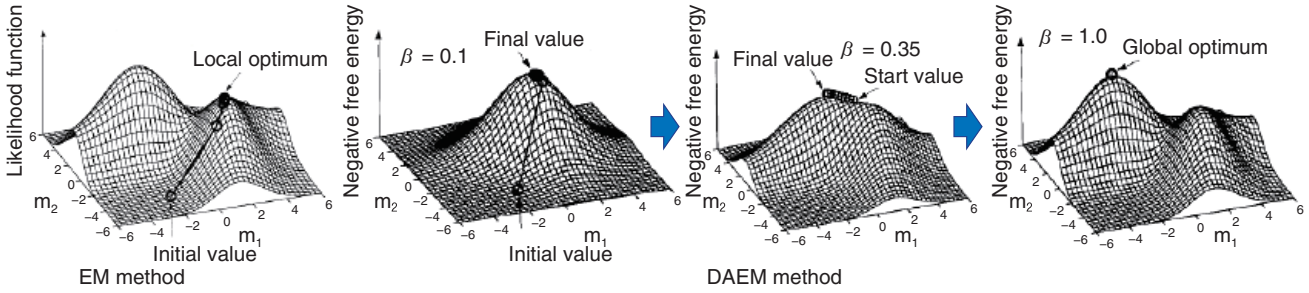
Fig. 2. Spatio-temporal multidimensional collective data analysis.

Kyoto, which generated a demand for young researchers in all of those laboratories. I was initially involved in basic research in the field of computer vision focusing on the theme of image analysis from line drawings, but my main interest at that time was statistical learning theory. This research of image processing that I had undertaken focused on *process*, that is, a method of cleaning up a picture or clarifying its contour. However, what I really wanted to know was not the shape of the picture but what the target object actually was. In other words, I was more concerned with achieving "understanding" or "recognition." To this end, I resolved to move to the field of machine learning, which I thought would provide me with more theoretical opportunities, and I applied for a transfer with my research plan in hand.

Machine learning began in the 1970s as a field within AI, and neural networks boomed in the 1980s as a type of machine learning applicable to a variety of application fields. This boom, however, lasted only a few years, and perhaps as a reaction to this short-lived technology, machine learning went on to evolve into somewhat low-key mathematical research called learning theory. Against this background, I took up research at NTT Communication Science Laboratories, but two years later, a turning point happened in my career; my research in the field of vector quantization resulted in awards that put me on the path to overseas research stays.

Then, having launched a new research theme, I took up the challenge of submitting a paper to Neural Information Processing Systems (NIPS), a prestigious conference in machine learning known to be highly competitive. That paper was the first one by a Japanese researcher to be accepted, and it provided me with an opportunity to present my research results (**Fig. 3**) [2]. I think my motivation behind this challenge was partially in response to the criticism I received from the then laboratory director, who had said to me, "Your group is doing nothing meaningful or useful."

Then, in the aftermath of the NIPS paper, I received an invitation from Dr. Geoffrey Hinton, a world authority on neural networks, to visit the University of Toronto in Canada, and I ended up spending about a month there. Dr. Hinton told me he didn't really care for statistics and mathematics very much, and contrary to my expectations, I did not receive direct research guidance from him. But thinking that I had to produce some results, I became absorbed in my work. Indeed, I was in crisis mode, as I had to accept the fact that my connection with Dr. Hinton would



Comparison of expectation-maximization (EM) method (conventional method) and the deterministic annealing EM (DAEM) method with respect to the likelihood function. With the EM method, the function converges to a local optimum near the initial value (left mesh graph), but with the DAEM method, the EM process is reformulated as the problem of minimizing the negative free energy dependent on a temperature parameter (β) so that the likelihood function converges to the global optimum (right mesh graphs).

Fig. 3. Overview of method reported in NIPS paper.

probably come to an end if I could not generate any results in that one month. I dived into my research forgetting to eat and sleep in an attempt to produce something one way or another, and in the end, the results that I achieved became the seeds of research that was again selected for publication by NIPS.

At that time, I wasn't sure what the results themselves would be useful for, but after Dr. Hinton evaluated them again, I was invited to a new research laboratory at University College London in the United Kingdom that he had founded, and I remember being called the "first visitor" in a lighthearted manner.

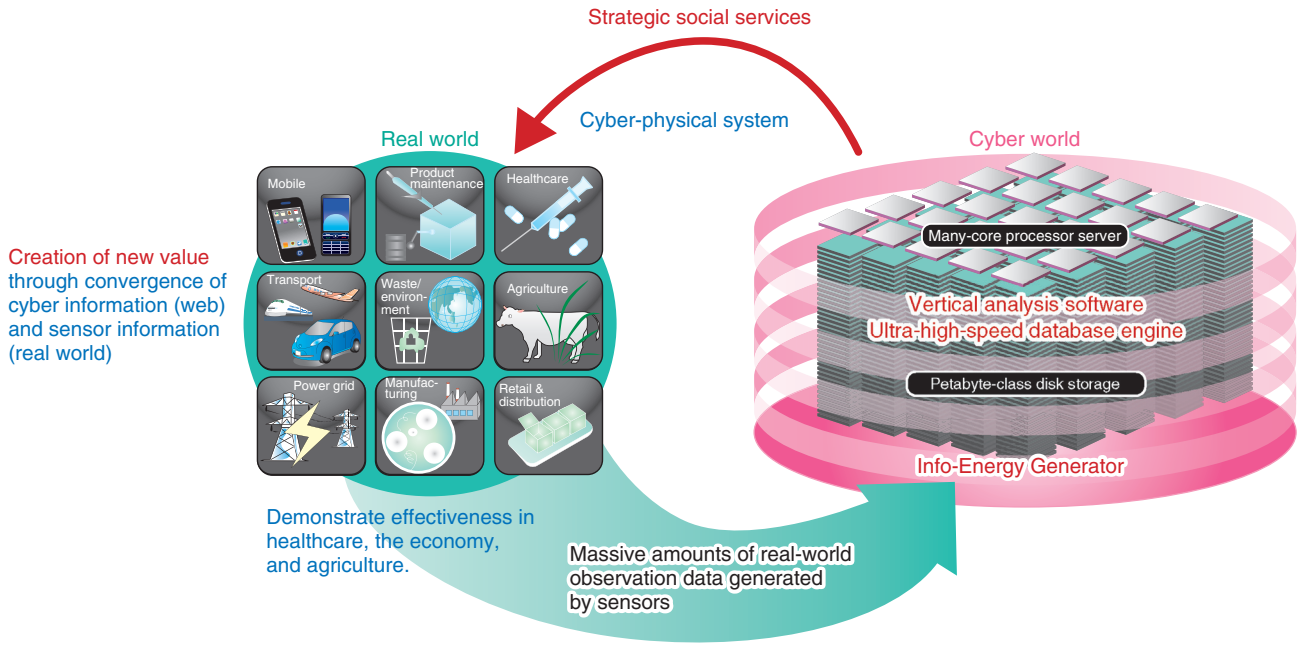
I then expanded my research into the field of speech recognition and took up technology for achieving a simpler and more efficient speech recognition model. This research was a joint effort with speech recognition researchers at that research laboratory, but the results achieved were evaluated by a number of parties and received various awards. The core of these results, however, consisted of basic theoretical research. On reflection, I feel that the sum total of the research on theoretical systems that I carried out in my younger days laid the foundation for my present research. The rebellious spirit of my senior researchers and the harsh criticism of the laboratory director at that time, the challenges presented to me by my group leaders, the existence of NTT Communication Science Laboratories, and the environment made available to me by Dr. Hinton have all contributed to what I am today.

The ability to produce your research theme is the key
There is much joy in making the most of opportunities and delivering results

—So your encounter with a pioneering researcher and your presentation at a highly competitive international conference were filled with drama!

Because of these experiences, my research activities expanded greatly from what had been work in a somewhat narrow field. The second half of the 1990s saw a trend toward investing in machine learning, and the multiplex topic classification technology for text that I developed was put to use in the *goo* portal site operated by NTT Resonant for classifying news articles and performing other functions. Then in 2005, I met Dr. Masaru Kitsuregawa, an authority on data engineering, and became involved in the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST) of the Cabinet Office, Government of Japan, in the role of a sub-theme leader in Dr. Kitsuregawa's research project.

In this program, 30 core researchers were selected from all research fields in Japan, and Dr. Kitsuregawa was the sole researcher selected from the field of information engineering. His project theme was the development of an ultra-high-speed database, and I was entrusted with the sub-theme of determining in what fields such a database, if achieved, could be applied (**Fig. 4**) [3]. Together with more than ten co-researchers, I focused on big data analysis in the field of healthcare. In particular, I took up the challenge of automatically recognizing nursing actions and



The project aims to achieve a prototype cyber-physical system in healthcare, the economy, and agriculture. This system will process big data collected in the physical world using an ultra-high-speed database, analyze the data using advanced machine learning technology, and feed back the results to society and implement them in social services [3].

Fig. 4. Overview of Dr. Kitsuregawa's project.

analyzing the effects of those actions on patients, working with doctors at Kyushu University Hospital.

Achieving automatic recognition of nursing actions was a very difficult endeavor. Although broad, obvious actions such as running could be easily recognized, relatively calm and understated actions such as patient interviewing and drawing blood presented a problem. That is, under the constraints of automatic recognition using just a few acceleration sensors affixed to a nurse so as not to hamper daily nursing tasks, the application of existing machine learning techniques could not achieve sufficient accuracy here for practical implementation. Doctors at the hospital would make disparaging comments like "Is this the best machine learning can do?" Additionally, as it appeared that no journal paper was in the offering under these circumstances, the collaborating doctors distanced themselves from the research one after another.

Nevertheless, as I was the one responsible for this research, I persisted with this theme for another year and a half. In the end, I succeeded in creating a meta-learning method that I used to perfect a high-accuracy automatic recognition technique. I then went on to analyze data covering about 9 million nursing actions and to analyze the effects of nursing intervention on the admitting and releasing of patients. This big data analysis was the first of its kind in the world, and it received high marks in the final FIRST evaluation for opening up a new path in healthcare through the use of information and communication technology.

More recently, I have come to be involved in research related to cosmic physics. This research involves using the Subaru Telescope on the summit of Mauna Kea on the island of Hawaii and is aimed at estimating and analyzing the distance to supernovae about 10 billion light years away and constructing a history of universe expansion. The researchers here need to instantly and automatically identify supernovae and collect data on them, and I have helped meet this need using machine learning technology.

Further classification of supernovae types is just getting under way. At the University of Tokyo, the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU) [4]—the client of this joint research—received praise for this new research approach using machine learning and consequently became the first Japanese research institution to get observation time on the Hubble Space Telescope, which has been providing valuable astronomical data

of historic proportions. In this way, Kavli IPMU has been able to obtain even more data on supernovae, for which I am grateful.

This means that we can now perform world-leading research by combining the wide-field Subaru Telescope and the high-accuracy Hubble Space Telescope, which is a very satisfying outcome.

—Surely you have gone where no one has gone before!

I would not make such a claim—I think it's more an outcome of good fortune. However, I would say that I have been successful in finding promising new themes and taking up the challenge of trying to solve difficult research and analysis problems. This process begins with the people one encounters and what kind of joint research one pursues, and looking back, I believe that making an effort to produce results and not simply giving up is how I have been able to create new opportunities.

For example, when given the chance to analyze nursing actions from Dr. Kitsuregawa, I certainly did not want to waste that valuable opportunity, and knowing that there would not be another chance if I produced nothing of merit, I felt a sense of urgency. I refused to give up and went on to create the meta-learning method, working one step at a time in a low-key manner for about two years.

Remembering the sense of excitement and reading as many journal papers as possible with great intent—steady effort opens up the future

—What would you say to young researchers?

To each of our young researchers, I would say, "Charge ahead in your chosen theme with all of your heart and soul." Many things in life like something someone says or current social trends can weaken your resolve, but don't lose focus—hold steady and be true to yourself. Don't become demoralized or run away in the face of harsh comments. At the same time, don't let the burden of responsibility make you continuously anxious; face the problem with a sense of balance. When the crunch comes, you may forget to sleep and eat as you look for a solution, but it's okay to take a break too.

In basic research, the probability of a hit is 1 out of 1000 in extreme cases, so a lot of time and patience are needed. Whether one can endure such frustration is the key to success. A superficial paper, though easy to write, cannot be very meaningful or influential. Like all of us, I have taken on difficult problems that I worked tirelessly on with my share of ups and downs. When I first took up the challenge of submitting a paper to NIPS and received the mail notifying me of its acceptance, I wanted to cry out in a loud voice from the bottom of my heart. I still can't forget that feeling of joy, and I ask all of you as well not to forget what such a sense of accomplishment feels like.

—What then, should researchers focus on when they are young?

At present, I'm involved in applied research making use of my research achievements to date. I am convinced that cultivating basic technology in a dedicated manner when I was young was extremely important in making this happen. I believe that the basics play an important role not just in research overall but in the life of the individual researcher too. For example, a technology called reinforcement learning that was first devised about 30 years ago is used in the AlphaGo *go* playing program that has recently become a topic of conversation. Back then, no one thought that this technology would be used in a program for playing go, but as it turned out, AlphaGo could not have been achieved without it. So I would like young researchers who have much time ahead of them to work at building such a foundation. Without accumulating basic technologies, nothing new will come into view.

—So accumulating results is important for young researchers having much time ahead of them.

Yes, and in addition, journal papers are useful in establishing a network of contacts. In my younger days, publishing papers was a motivating force behind my research, but nowadays, I pursue research that, if anything, can have a direct impact on society. To this end, I feel that encountering brilliant researchers to form partnerships for collaborative work is essential, and to get such topnotch researchers interested in your work, it is important, after all, to focus intently on writing good papers and getting noticed. Compared with the past, it is easy to search for papers in today's environment, which makes it easier to obtain evaluations from outside the laboratory than from inside.

It is also important to hold discussions with a variety of researchers to generate new ideas. Each researcher has a problem that he or she is dealing

with, and simply talking about your respective problems and worries may provide a hint to solving your own research problem. Here, getting your partner to open up requires that you build a trustworthy relationship, so when there are partners in dissimilar fields, I make it a point to visit with those partners often. I myself am an elemental-technology techie, that is, a "hammer-and-saw" tool techie! And to provide tools with value, it is important to talk beforehand with a specialist in that area on what kind of house we should construct. There is no need to worry about each other's status here—we should form relationships frequently to learn new things from others.

"What" is what's necessary for research, while "how" is what's important in a paper—top-level researchers focus on "what"

—Dr. Ueda, in closing, what do you see as future issues in machine learning, and what is your personal outlook for the future?

I place much importance on educating young researchers. Since top-level young researchers study an awful lot on their own, I'm always thinking of how I can broaden their research fields and provide them with a stimulating environment rather than simply guiding them in their research. I want to create a space in which they can demonstrate their abilities freely. By the way, machine learning, while heretofore having a somewhat shunned existence, now has wind in its sails, so to speak, as AI comes to be generally accepted. However, this boom too will one day subside, so I want to establish an environment in which young researchers can continue their research at that time. This is the mission of a senior researcher.

I would also like to contribute to solving problems that researchers in a variety of fields are dealing with. At this point in my research career, I feel that there is more value in making concrete contributions than in writing papers. For example, I want to see ambient intelligence and spatio-temporal prediction that I talked about earlier to continue developing as far as verification experiments and actual operation. This type of research on social infrastructure cannot proceed by one company alone. Partnerships must be formed with related companies, and progress must be steadily made while making adjustments.

Of course, this is easy to say but difficult to achieve in practice. If any party pursues only its own interests, the collaborative effort will break down, so a balance between cooperation and competition must be achieved. Writing research papers is not the only work of a researcher. I believe that experienced researchers and research leaders are required to look at things from their partners' point of view and to achieve even better technology.

Although "how" is of course important in the case of a paper, "what" is important in the activities of an experienced researcher. A specialist is interested in "how," but the impact of technology on society is "what." Research that can clearly explain what has been developed and what purpose it can serve in society is research at its best. Actually, it's only recently that I became aware of this. This is because elemental-technology techies are a species of researchers who are often interested only in "how." In addition, a small amount of difference in "how" has nearly no value. Furthermore, if a technology is not usable in many areas, that technology can be only halfway satisfying. At the risk of repeating myself, if a researcher cannot produce results in the form of "what," that research will likely have no impact on society.

Services and products such as Uber, Pokémon Go, and AlphaGo that have swept society are all combinations of existing technologies; they have not made any contributions in the form of new technologies. However, beating the world's champion go player even with a combination of existing technologies is an outstanding achievement that can have a great impact. If researchers themselves in an information society dominated by the service industry can produce value and high-impact results, certainly the quality of research will change significantly.

Today's young researchers worry that their papers may go unnoticed even if the papers are well written. For this reason, I would like to provide extensive guidance to young researchers on the ability to produce value and results. I point out one more time that communicating with researchers in a variety of fields is a great way of refining one's ability to produce results as a researcher.

References

[1] F. Naya, M. Miyamoto, and N. Ueda, "Optimal Crowd Navigation via Spatio-temporal Multidimensional Collective Data Analysis," NTT Technical Review, Vol. 15, No. 9, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201709fa5.html

[2] N. Ueda and R. Nakano, "Deterministic Annealing Variant of the EM Algorithm," Neural Information Processing Systems 7 (NIPS7), MIT Press, Cambridge, MA, USA, pp. 545–552, 1995.

[3] Website of the "Development of the Fastest Database Engine for the Era of Very Large Database and Experiment and Evaluation of Strategic Social Services Enabled by the Database Engine" project. http://www.tkl.iis.u-tokyo.ac.jp/FIRST/outline.en.html

[4] Website of Kavli IPMU, http://www.ipmu.jp/en/news-information

Trademark notes

■ Interviewee profile

Naonori Ueda

NTT Fellow, Head of Ueda Research Laboratory and Director of Machine Learning and Data Science Center, NTT Communication Science Laboratories.

He received a B.S., M.S., and Ph.D. in communication engineering from Osaka University in 1982, 1984, and 1992. In 1984, he joined NTT Electrical Communication Laboratories, where he was engaged in research on image processing, pattern recognition, and computer vision. In 1991, he joined NTT Communication Science Laboratories. From 1993 to 1994, he was a visiting scholar at Purdue University, Indiana, USA. He was a director of NTT Communication Science Laboratories (April 2010–March 2013). In addition to his current positions as head of Ueda Research Laboratory (NTT Fellow) and director of the Machine Learning and Data Science Center, he also serves as deputy director of the RIKEN Center for Advanced Intelligence Project, established in April 2016. He is also a visiting professor at the Graduate School of Informatics, Kyoto University, and at National Institute of Informatics.

Basic Research in the Era of Artificial Intelligence, Internet of Things, and Big Data—New Research Design through the Convergence of Science and Engineering

Eisaku Maeda

Abstract

To build a productive relationship between humans and artificial intelligence (AI), we must grasp the current situation as accurately as possible and make investments in the future toward developing such a relationship. This article introduces how we see and interpret the AI, Internet of Things, and big data era from the standpoint of promoting research and development of basic technologies.

Keywords: artificial intelligence, Internet of Things, big data

1. Artificial intelligence that approaches and gets closer to people

In the Future of Go Summit held in May 2017, the artificial intelligence (AI) program AlphaGo competed against and defeated China's 18-year old Ke Jie, the world's number one ranked *go* player. AlphaGo produced three solid, consecutive wins against Ke to win the best-of-three match. Computer AI programs had already defeated humans in chess and *shogi* (Japanese chess); however, because go is more complex than chess or shogi, it was said that it would take more than 10 years to develop an AI program that could defeat go professionals. However, the calculation abilities of AlphaGo quickly exceeded human capabilities.

This event in the world of games provides a suitable lesson for us to think about for the future. A new partner has come into our lives, and from now on we will have to share our lives with such partners. This may seem bothersome at first, but, whether we like it or not, we now have to put our heads together and find ways of building appropriate relationships with these new AI devices.

2. A new future emerging from convergence of AI, Internet of Things, and big data

The convergence of three fundamental elements—new data technology known collectively as AI, new data infrastructure called the Internet of Things (IoT), and vast collections of data of unprecedented size and quality known as big data—is now having a major transformative impact on the world [1]. Yet today we can only foresee a tiny fraction of the services and technologies that might arise in the years ahead, so we must rack our collective brains to come up with possible applications and to consider the potential impacts on society. Convergence of AI × IoT × big data will open the way to global-scale technological infrastructure in two basic ways: digitization of information permitting precise replication of data, and global networking that enables simple accumulation and dissemination of data (**Fig. 1**). As we build on this infrastructure, we are poised to create a new future with machine learning and algorithms for processing

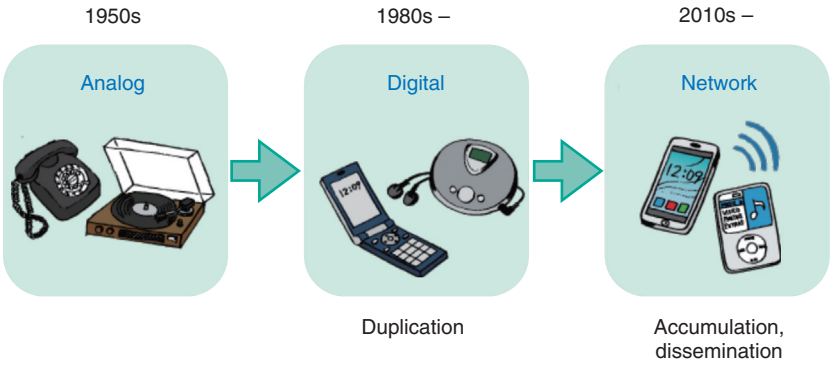


Fig. 1. Evolution of technology.

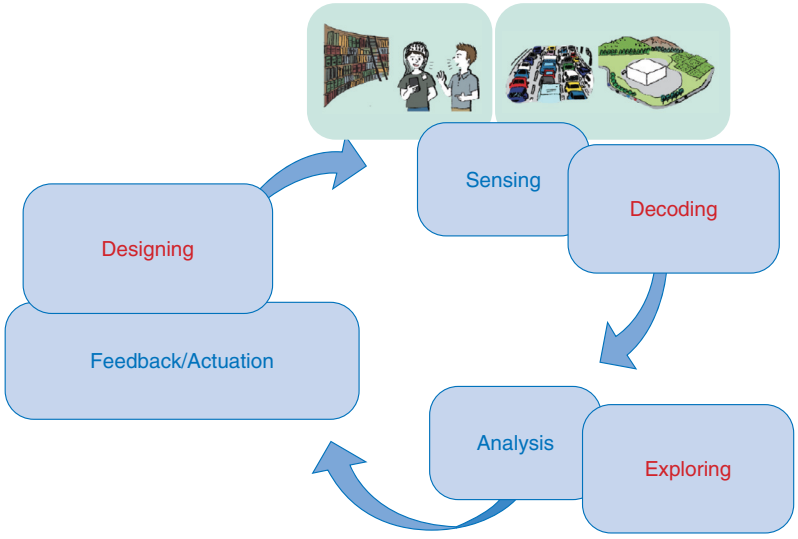


Fig. 2. Decode, explore, and design.

information.

3. Cycle of decoding, exploring, and designing

Data encompassing everything under the sun—people, things, environments—are acquired, decoded, explored, then designed and fed back to society (**Fig. 2**) [2]. Intelligence is modeled from every conceivable type of data related to real people, for example, writings, actions, speech, images, sounds, biological reactions, muscle activity, and brain activity. As this process unfolds, the day is fast approaching when intelligence will be able to be broken up and reassembled as components that can be treated as products on the open market.

Pure science and engineering have always been pursued as independent disciplines, but this is now undergoing a major change caused largely by the developments outlined above. With the availability of low-cost sensing and intelligence, we are now able to amass vast amounts of basic scientific data across a wide range of fields within a remarkably short period of time, which can readily be applied to engineering tasks thanks to more reliable processing and analysis.

At the same time, however, this is creating a need for people who are skilled in reading raw data, designing experimental plans, and performing tasks in other empirical sciences in engineering fields such as signal processing and data analysis. The former might be called *engineering of science*, while the

Fig. 3. Navigating real-world data.

latter is the *science of engineering*. Feedback loops from science to engineering to implementation and back to science repeat to become ever thicker, and the sequence revolves at an ever faster rate. Thus, even areas of scientific research that up to now have been conducted in idyllic independence must be strategically reexamined in terms of overall economic efficiency.

4. Who steers the real world

The AI boom will eventually lose momentum, but even at a slower pace, it will continue to make steady progress [3, 4]. What ultimate vision can we anticipate in the future? Just as Marvin Minsky, the father of AI, sought to peer inside the brain to unravel the secret of intelligence, it should soon be possible for ordinary non-specialists to easily access and view the world's data. People will be able to view and manipulate and control a full range of decoded data going well beyond mere sounds and photographic images. Even today we see young children, only two and three years old, who are surprisingly adept at playing with tablet computers and watching YouTube videos and other content. In much the same way, we can fairly assume that you and I—ordinary people—will be able to navigate real-world data with nothing more than a tablet computer in the not-too-distant future (**Fig. 3**). Whether this is good or bad, it certainly suggests that we are moving toward a more open democratization of information.

5. Communication science that enables corevo®

The NTT Group recently rolled out a new AI brand called corevo®. We are proud to see corevo on the market, for this new technology is largely based on basic research from our labs, NTT Communication Science Laboratories. The Feature Articles in this issue introduce some of our latest findings and research results [5–9].

References

[1] E. Maeda, "From Information Transmission to Mutual Understanding: Paradigm Shift in the Age of Data," NTT Technical Review, Vol. 14, No. 11, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201611fa1.html
[2] E. Maeda, "Embracing Information Science and Technology—Decoding, Exploring, and Designing the World," NTT Technical Review, Vol. 13, No. 11, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201511fa1.html
[3] E. Maeda, "The Evolution of Basic Research," NTT Technical Review, Vol. 12, No. 11, 2014.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201411fa1.html
[4] E. Maeda, "Advanced Research in Speech, Language, and Hearing for Communication of the Future," NTT Technical Review, Vol. 11, No. 12, 2013.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201312fa1.html
[5] T. Kaneko, K. Hiramatsu, and K. Kashino, "Generative Personal Assistance with Audio and Visual Examples," NTT Technical Review, Vol. 15, No. 11, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201711fa2.html
[6] M. Nishino, N. Yasuda, S. Minato, and M. Nagata, "Efficient Algorithm for Enumerating All Solutions to an Exact Cover Problem," NTT Technical Review, Vol. 15, No. 11, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201711fa3.html
[7] J. Suzuki and M. Nagata, "Memory-efficient Word Embedding Vectors," NTT Technical Review, Vol. 15, No. 11, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201711fa4.html
[8] T. Fukiage, T. Kawabe, and S. Nishida, "Synthesizing Ghost-free Stereoscopic Images for Viewers without 3D Glasses," NTT Technical Review, Vol. 15, No. 11, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201711fa5.html
[9] M. Delcroix, K. Kinoshita, A. Ogawa, S. Karita, T. Higuchi, and T. Nakatani, "Personalizing Your Speech Interface with Context Adaptive Deep Neural Networks," NTT Technical Review, Vol. 15, No. 11, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201711fa6.html

Trademark notes

Eisaku Maeda
Former Vice President, Head of NTT Communication Science Laboratories*.

He received a B.E. and M.E. in biological science and a Ph.D. in mathematical engineering from the University of Tokyo in 1984, 1986, and 1993. He joined NTT in 1986. He was a guest researcher at the University of Cambridge, UK, in 1996–1997. His research interests are statistical machine learning, intelligence integration, and bioinformatics. He was awarded the Information Processing Society of Japan (IPSJ)'s 45th anniversary best paper on the next 50 years of information science and technology for his paper "Resurgence of Fairies and Goblins—A Proposal for the New Vision of "Ambient Intelligence—." He is a fellow of IEICE (Institute of Electronics, Information and Communication Engineers of Japan) and a senior member of IEEE (Institute of Electrical and Electronics Engineers) and IPSJ.

*Until the end of August, he was Vice President and head of NTT Communication Science Laboratories. He is currently a professor at Tokyo Denki University.

Generative Personal Assistance with Audio and Visual Examples

Takuhiro Kaneko, Kaoru Hiramatsu, and Kunio Kashino

Abstract

The rapid progress of deep learning is affecting the world we live in. Media generation (i.e., image and audio generation) is a typical example of this progress, and impressive research results are being reported around the world. In this article, we first overview this very active research field. Then we introduce our efforts in developing a generative personal assistance system with audio and visual examples. Specifically, we explain how our new deep-learning approach will overcome the limitations encountered in existing studies on personal assistance systems. Finally, we discuss future directions in the media generation field.

Keywords: deep learning, media generation, personal assistance

1. Rapid progress in deep learning

The rapid progress achieved recently in deep learning is having an impact on the world we live in. Media generation (i.e., image and audio generation) is a typical example of this progress. Many studies are being done in this field, and amazing research results are being reported around the world.

For example, StackGAN [1] can automatically synthesize photorealistic images only from a text description. Neural style transfer [2] can convert arbitrary images to arbitrary-style ones (e.g., Gogh-style or Monet-style images) without any manual processing.

Until a few years ago, the main objective in deep learning was to improve accuracy for comparatively easily defined tasks such as image classification and speech recognition. However, more complex tasks have recently been tackled by introducing new models and theory. Also, a major breakthrough was recently achieved in the media generation field due to the emergence of deep generative models. Consequently, expectations for media generation are rising as a tool to embody various wishes.

2. Generative personal assistance

When we do new things (e.g., try to throw a ball faster, try to pronounce English more fluently, or try to acquire an appropriate facial expression according to the time, place, or occasion), it is sometimes difficult to know what the most effective methods for doing so are, and this can cause frustration. One conceivable solution is to find a teacher and have him or her teach us; another is to search via the Internet or books by oneself.

The former approach is useful because it provides detailed instructions, but it is not always easy to find a suitable teacher. The latter approach is useful because it does not require help from anyone else, but it is not easy to find the optimal solution to fit the individual. To solve this problem, we aim to develop a generative personal assistance system as a means of learning new things. It is advantageous in that it offers individuality and concreteness provided in the former approach above, as well as automaticity provided in the latter approach.

3. Generative personal assistance with audio and visual examples

The concept of our proposed generative personal assistance system providing audio and visual examples is illustrated in **Fig. 1**. We achieve individuality by analyzing the relevant data based on the data

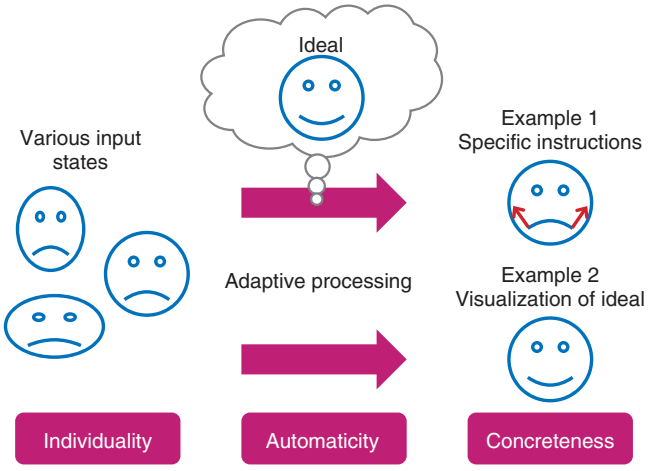


Fig. 1. Generative personal assistance with audio and visual examples (e.g., facial expression improvement).

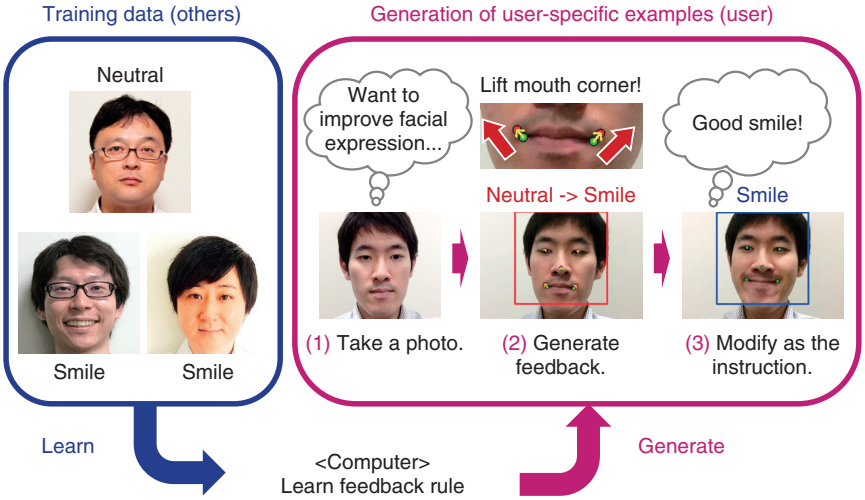


Fig. 2. Operation example of proposed feedback system.

provided by the user (i.e., images and audio). We achieve concreteness by visualizing or auralizing the specific instructions or the ideal state. We achieve automaticity by ensuring the above process is conducted automatically.

An operation example of the system [3] developed to provide visual feedback to someone who wants to improve their facial expression is shown in **Fig. 2**. In this system, we learn the feedback rule using the pre-prepared data. We assume that the data are collected not from the actual user but from the general public. Moreover, we assume that we can only collect one-shot data (i.e., one facial expression) for one person. This condition enables us to achieve automaticity without requiring a large data collection effort.

In generating feedback, we achieve individuality by analyzing and calculating feedback based on the user-provided image. We then display improvement guidelines with arrows. This feedback is concrete and interpretable, enabling the user to achieve the target state by following the instructions. This example targets facial expression improvement, but the same idea can be applied to other tasks (e.g., improving pitching in baseball or improving pronunciation).

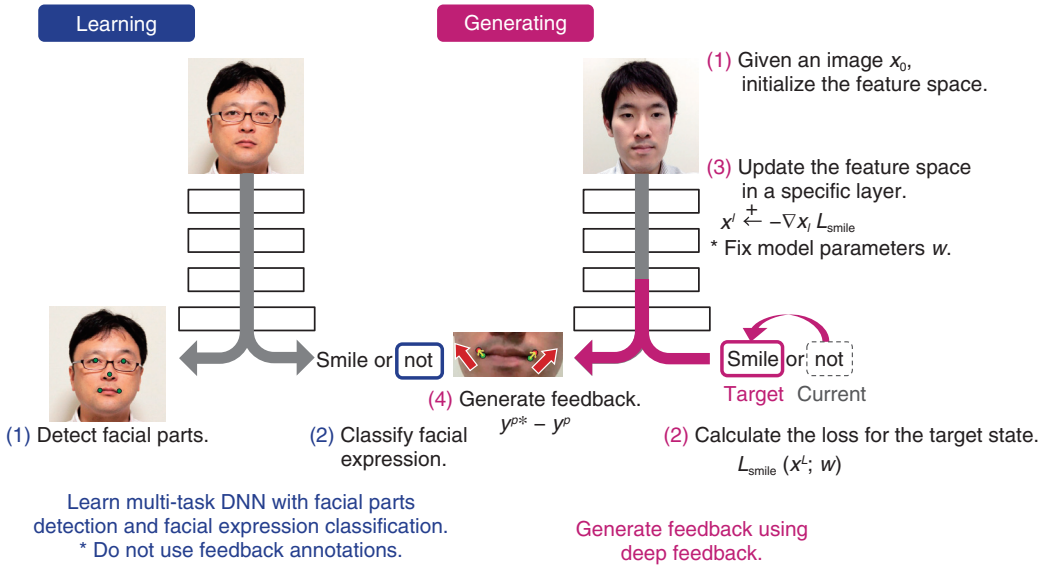


Fig. 3. Feedback generation with deep feedback.

4. Input variation and output concreteness

Two challenges need to be overcome in order to achieve this system. The first is how to apply it to various input data. This system is aimed for use by a diverse range of people with different ages, genders, and races. Therefore, it is necessary to adapt the system to provide the optimal feedback for each person. The main approach used in previous feedback systems is rule-based, in which the feedback rule is defined manually. This imposes a large rule-creation cost when there is a large amount of variation data. We propose to solve this problem by using a learning-based approach.

The second challenge is not only to recognize the current state but also to generate detailed feedback based on it. Previous systems also used a learning-based approach that can be applied to various inputs, but they can only recognize the current state; for example, they can only classify facial expressions, estimate the degree of a smile, and detect the different parts of the face. It is not easy to extend this approach to generate feedback because the typical learning scheme requires the pair data of input and output, that is, the user image and correct feedback in this case. In general, feedback annotations require professional knowledge. Therefore, it is not easy to collect a large amount of data. This limitation makes it difficult to develop a feedback system using a learning-based approach.

5. Feedback generation with deep feedback

To overcome these two challenges, we propose a new deep neural network (DNN)-based method that can learn and generate feedback without the annotations of correct feedback. The process flow of this method is shown in **Fig. 3**. In the learning step, we train a multitask DNN in facial parts detection and facial expression classification. This model also requires annotation data—that is, facial expression annotations and facial parts annotations—for training, but they are easier to collect than feedback annotation because they do not require professional knowledge. In the resulting model, the feature space has relevance to input images, facial parts, and facial expression classes that are key factors for generating feedback.

In the generating step, we apply a novel propagation method called deep feedback to extract the feedback information from the feature space. (1) Given an image, the system initializes the feature space and estimates the current position of facial parts and the facial expression class. (2) For the classification output, the system calculates the loss of the target state. (3) The system reduces this loss by optimizing the feature space in a specific layer using back-propagation. In the experiment, we found nonlinear transformation through deep back-propagation is useful for adapting feedback to various inputs. Note that in general training the model parameters are updated while fixing the

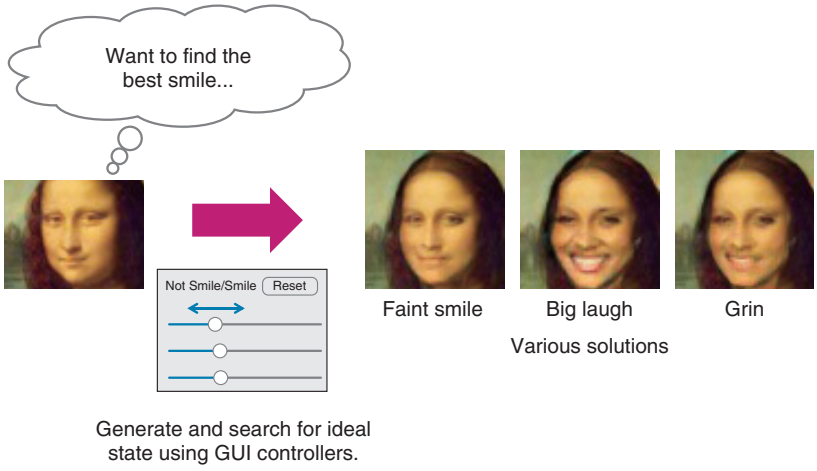


Fig. 4. Interactive operations for generating and searching for ideal state.

feature space, but in the deep feedback process the feature space is updated while fixing the model parameters. (4) The optimal facial parts position is calculated based on the optimized feature space using forward propagation. Finally, arrows indicating the difference between the current and optimal facial parts position are displayed.

6. Interactive operations for generating and searching for ideal state

In the above framework, an algorithm finds and visualizes only one solution for smiling, but there are actually various solutions for smiling (e.g., a grin, faint smile, or big laugh). In such a case, it is not easy for a computer to select and generate an optimal one without the user's interaction. We were motivated by this to develop a novel system [4] where a user can generate and search for his/her ideal state interactively using typical GUI (graphical user interface) controllers such as radio buttons and slide bars. A conceptual image of this system is shown in **Fig. 4**.

7. Representation learning using deep attribute controller

To develop this system, we need to obtain a representation space that is disentangled, expressive, and controllable. First, attributes, for example, smiling, and identity need to be disentangled in the representation space in order to change the attributes independently from the identity. Second, the representation space needs to be expressive enough to represent the attribute variations. Third, controllability is important because our goal is to enable a user to intuitively control attributes. One of the challenges in learning such a representation space is to learn the space without a detailed description of attributes. This constraint is important because it is not easy to obtain a detailed description of attributes due to the complexity and difficulty in defining a rule for organizing them.

To solve this problem, we propose a novel method called deep attribute controller to learn disentangled, expressive, and controllable representations only from the binary indicator representing the presence or absence of the attribute. In particular, we propose a conditional filtered generative adversarial network (CFGAN), which is an extension of the generative adversarial network (GAN) [5]. The GAN is a framework for training a generator and discriminator in an adversarial (min-max) process.

The generator maps a noise to data space and is optimized to deceive the discriminator, while the discriminator is optimized to distinguish a generated and real sample and *not to be deceived* by the generator. The CFGAN incorporates a filtering architecture into the generator input, which associates an attribute binary indicator with a multidimensional latent variable, enabling the latent variations of the attribute to be represented. We also define the filtering architecture and training scheme considering controllability, enabling the attribute variations to be intuitively controlled. The CFGAN architecture is shown in **Fig. 5**.

The above explanation describes the case where the CFGAN is applied to learn smile variations, but the same scheme can be used to learn other attribute

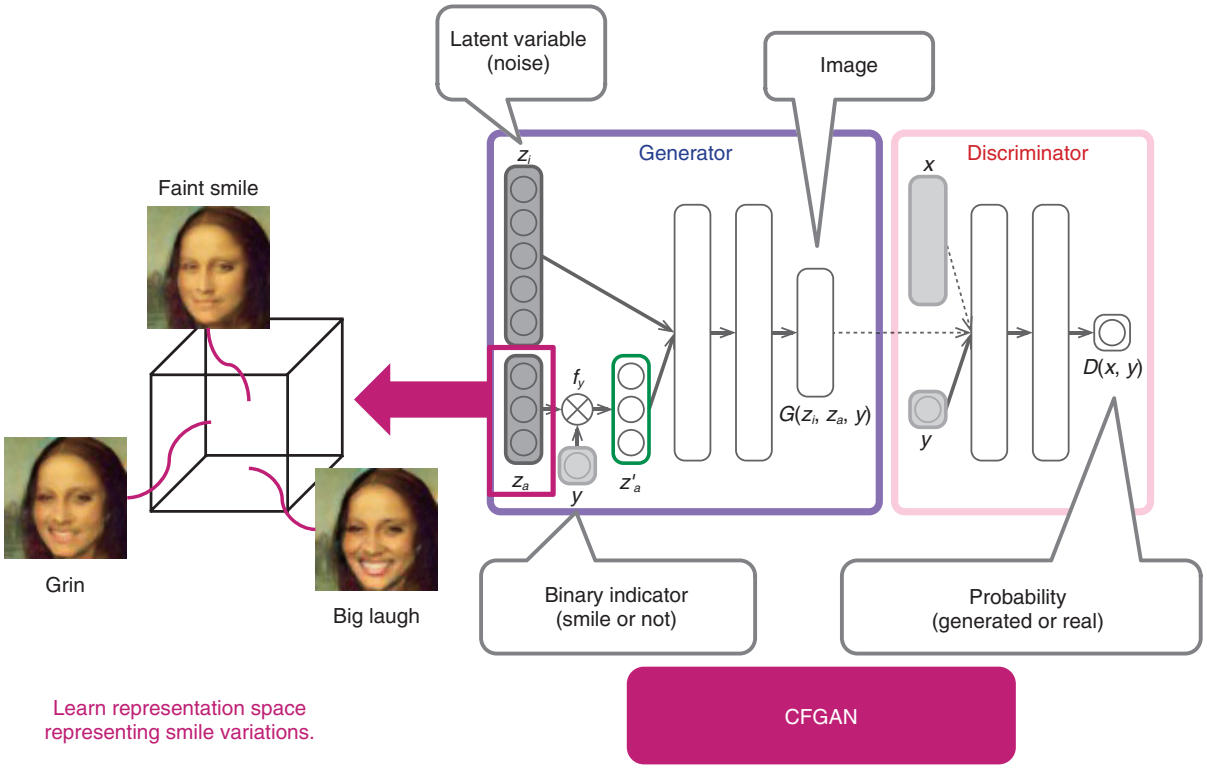


Fig. 5. Representation learning using deep attribute controller.

variations such as ages, hair styles, and genders. Moreover, it can be applied not only to images but also to audio data. For application to audio, we also developed some essential technologies consisting of realistic speech synthesis [6, 7] and realistic voice conversion [8] methods.

8. Future direction

The key objective of our approaches is to use media generation to develop an affinity with—that is, get close to—users. These technologies are essential for not only personal assistance but also for embodying users' wishes through concrete media information. In the future, we aim to establish the technology to generate exceedingly high quality media to meet any expectation. To achieve this, we are working to cultivate the imagination, knowledge, and experience of media generation.

References

[1] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks," Proc. of the 16th International Conference on Computer Vision (ICCV 2017), pp. 5907–5915, Venice, Italy, Oct. 2017.

[2] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," Proc. of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), pp. 2414–2423, Las Vegas, NV, USA, June/July 2016.

[3] T. Kaneko, K. Hiramatsu, and K. Kashino, "Adaptive Visual Feedback Generation for Facial Expression Improvement with Multi-task Deep Neural Networks," Proc. of the 24th ACM International Conference on Multimedia (ACMMM 2016), pp. 327–331, Amsterdam, Netherlands, Oct. 2016.

[4] T. Kaneko, K. Hiramatsu, and K. Kashino, "Generative Attribute Controller with Conditional Filtered Generative Adversarial Networks," Proc. of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), pp. 6089–6098, Honolulu, Hawaii, USA, July 2017.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," Proc. of Advances in Neural Information Processing Systems 27 (NIPS 2014), pp. 2672–2680, Montreal, Quebec, Canada, Dec. 2014.

[6] T. Kaneko, H. Kameoka, N. Hojo, Y. Ijima, K. Hiramatsu, and K. Kashino, "Generative Adversarial Network-based Postfilter for Statistical Parametric Speech Synthesis," Proc. of the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017), pp. 7910–4914, New Orleans, USA, Mar. 2017.

[7] T. Kaneko, S. Takaki, H. Kameoka, and J. Yamagishi, "Generative Adversarial Network-based Postfilter for STFT Spectrograms," Proc. of the 18th Annual Conference of the International Speech Communication Association (Interspeech 2017), pp. 3389–3393, Stockholm, Sweden, Aug. 2017.

[8] T. Kaneko, H. Kameoka, K. Hiramatsu, and K. Kashino, "Sequence-to-sequence Voice Conversion with Similarity Metric Learning Using

Generative Adversarial Networks," Proc. of 18th Annual Conference of the International Speech Communication Association (Interspeech 2017), pp. 1283–1287, Stockholm, Sweden, Aug. 2017.

Takuhiro Kaneko

Researcher, Recognition Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.E. and M.E. from the University of Tokyo in 2012 and 2014 and began Ph.D. studies at the University of Tokyo in 2017. He joined NTT Communication Science Laboratories in 2014, where he studies computer vision, image processing, speech processing, pattern recognition, and machine learning. His interests include image generation, speech synthesis, and voice conversion using deep generative models. He received the Hatakeyama Award from the Japan Society of Mechanical Engineers in 2012 and the ICPR2012 Best Student Paper Award at the 21st International Conference on Pattern Recognition in 2012. He received the Institute of Electronics, Information and Communication Engineers (IEICE) ISS (Information and Systems Society) Young Researcher's Award in Speech Field in 2017. He is a member of IEICE and the Information Processing Society of Japan (IPSJ).

Kaoru Hiramatsu

Senior Research Scientist, Supervisor, and Leader of Recognition Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.S. in electrical engineering and an M.S. in computer science from Keio University, Kanagawa, in 1994 and 1996, and a Ph.D. in informatics from Kyoto University in 2002. He joined NTT Communication Science Laboratories in 1996 and has been working on the Semantic Web, sensor networks, and media search technology. From 2003 to 2004, he was a visiting research scientist at the Maryland Information and Network Dynamics Laboratory, University of Maryland, USA. He is a member of IPSJ and the Japanese Society for Artificial Intelligence (JSAI).

Kunio Kashino

Senior Distinguished Researcher, Head of Media Information Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. from the University of Tokyo in 1995. He has been working on media information processing, media search, and cross-modal scene analysis. He is also an adjunct professor at the Graduate School of Information Science and Technology, the University of Tokyo, and a visiting professor at the National Institute of Informatics. He is a senior member of IEEE (Institute of Electrical and Electronic Engineers) and IEICE, and a member of the Association for Computing Machinery, IPSJ, the Acoustical Society of Japan, and JSAI.

Efficient Algorithm for Enumerating All Solutions to an Exact Cover Problem

Masaaki Nishino, Norihito Yasuda, Shin-ichi Minato, and Masaaki Nagata

Abstract

We introduce an algorithm that finds all solutions to an exact cover problem. Many real-world tasks including designing apartment layouts and electric circuits can be formulated and solved as exact cover problems. Our algorithm can solve exact cover problems up to 10,000 times faster than the previous method. Moreover, our method compresses and stores all solutions and so can efficiently find the solutions that satisfy several constraints. Therefore, our algorithm can efficiently find good solutions to exact cover problems found in the real world.

Keywords: algorithm, data structure, exact cover problem

1. Exact cover problems

Most people have had some experience with puzzles such as crosswords, Sudoku, and Rubik's cubes. Many computer science researchers have tried for decades to design algorithms* that can solve these puzzles by using computers. As a result, several efficient algorithms have been developed that can solve these puzzles much faster than humans can.

However, puzzles that are difficult for humans are also difficult for computers. Rubik's cube and Sudoku are known to be NP-complete problems, and they become drastically more difficult as the problem size increases. For example, if we increase the number of squares on each face of a Rubik's cube to 16, 25, 36,…, then solving the puzzle requires exponentially more time. This inherent difficulty of puzzles is why researchers are continuing to pursue more efficient algorithms that can solve large and complex puzzles.

We have developed a new algorithm that can efficiently find all the solutions of an exact cover problem, a fundamental problem in combinatorics. An example of an exact cover problem is to find a set of rows of a binary matrix X (a matrix whose elements are either 0 or 1), where the selected rows must contain exactly one numeral 1 in every column. An example of an exact cover problem is shown in **Fig. 1**. If the matrix shown in the figure is given as the input, then the set of rows (1, 3) has exactly one 1 in every column and is thus a solution to the exact cover problem. An exact cover problem may have multiple solutions. This example problem also has another solution (2, 3, 5). Our algorithm can find all of the solutions to this exact cover problem.

Our algorithm can also be used to find all the solutions to puzzles that can be formulated as exact cover problems. A polyomino is a puzzle that involves arranging tiles on a board using a set of pieces consisting of square cells, where all pieces are used and each cell on the board is covered by square cells, with no spaces left uncovered and no pieces overlapping. This is a typical example of a puzzle that can be formulated and solved as an exact cover problem.

The example in **Fig. 2** shows how tetromino puzzles, a kind of polyomino where every piece is made

* Algorithm: A computation procedure for solving a problem. A computer can solve various problems by running algorithms.

	A	B	C	D	E	F
1	1	1	1	0	1	0
2	1	1	0	0	0	0
3	0	0	0	1	0	1
4	0	0	1	1	0	1
5	0	0	1	0	1	0

The set of rows (1, 3) is a solution to the problem represented by the matrix.

Fig. 1. An exact cover problem.

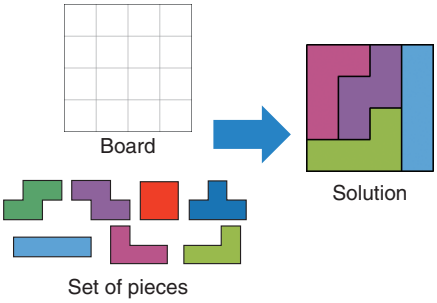


Fig. 2. Example of polyomino (tetromino) puzzle.

by connecting four square cells, can be formulated as an exact cover problem. Every column of the problem matrix corresponds to a cell of the 4×4 input board (thus, there are 16 columns), and every row corresponds to an arrangement of a piece. If the cell corresponding to the j-th column is covered by the arrangement of a piece corresponding to the i-th row, we set matrix element x_{ij} to 1. Otherwise, we set x_{ij} to 0. A solution to the exact cover problem formulated in this way represents a set of arrangements of pieces covering all of the cells on the board with no overlapping pieces. Hence, it is a solution to the tetromino puzzle.

Other than puzzles, several real-world problems can be formulated and solved as exact cover problems. For example, designing the layout of an apartment can be regarded as solving a polyomino puzzle, where every piece corresponds to a room. The problem of designing the layout of an electric circuit can also be formulated and solved as an exact cover problem that is similar to a polyomino puzzle. Hence, our algorithm can find good solutions to these problems.

2. Algorithm for solving exact cover problems

Exact cover problems are known to be NP-complete. Donald E. Knuth's algorithm DLX is accepted as a state-of-the-art algorithm for finding all solutions to an exact cover problem [1]. DLX solves a problem by performing an exhaustive search. Given input binary matrix X, DLX selects the first row and checks whether or not it is a solution to the exact cover problem. If it is not a solution, then it selects the pair of the first and second rows and checks whether or not that combination is a solution. DLX repeatedly adds rows to the current set of rows and checks whether it is a solution. If adding the i-th row to the current set makes more than two 1's appear in a column, then it cannot be a solution. Therefore, DLX removes the most recently added row from the set and adds a different row to the set to continue the search procedure. In this way, algorithm DLX finds all solutions by repeatedly adding and deleting rows to the set of rows to check all possible combinations of rows. Although the exhaustive search procedure is straightforward, DLX accelerates the search by exploiting a specialized data structure.

DLX can efficiently find all solutions to an exact cover problem if the number of solutions is limited. However, since it is an exhaustive search method, it takes an excessive amount of time if there are many solutions. Real-world exact cover problems sometimes have a huge number of solutions. For example, small polyomino puzzles can have more than one billion solutions.

Our new algorithm improves DLX to achieve practical speeds even when there are huge numbers of solutions [2]. The key idea of the proposed algorithm is to store all the solutions found in a search. If an exact cover problem has many solutions, then it is highly likely that they will be similar in several ways. For example, there are many solutions to a polyomino puzzle that has the same placement of pieces on the left half of the input board. If we memorize all the solutions, they can be used as hints to find similar solutions and thus accelerate the search procedure.

Although storing all the solutions may increase the search speed, memorizing billions of solutions in a naïve manner is unrealistic. Our algorithm uses the data structure called a zero-suppressed binary decision diagram (ZDD) to store the set of found solutions. A ZDD represents the set of solutions as a directed graph. We show in **Fig. 3(b)** an example ZDD that represents the set of two solutions (**Fig. 3(a)**) of the tetromino puzzle in Fig. 2. This ZDD has two paths that start from the root node and end at terminal node ⊤. These two paths correspond to the two solutions. Since a ZDD shares partial paths, it yields a

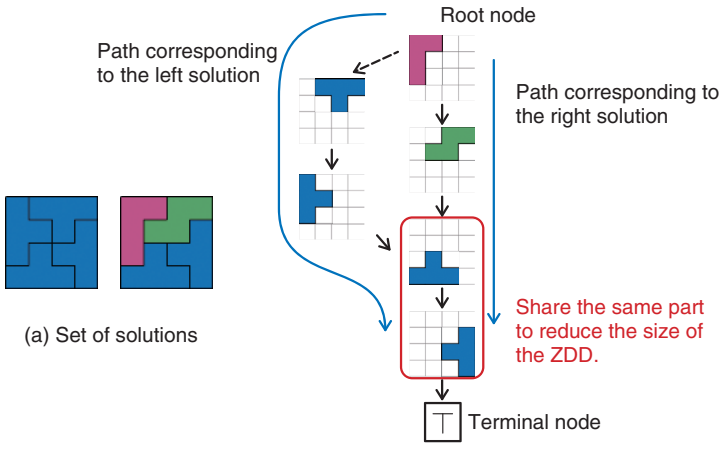


Fig. 3. ZDD representing a set of solutions to an exact cover problem.

small graph that can represent huge numbers of solutions. The ZDD in the figure shares two common edges of paths in order to reduce the size of the graph.

For example, a tetromino tiling problem with a board size of 10×10 has 7,213,560,548,906,621 solutions. If we use the naïve representation, the computer would require 100 petabytes (10^{18} bytes) of memory. In contrast, with our ZDD approach, the set is represented as a directed graph with 16,476,396 nodes. Such a ZDD requires only 300 megabytes (10^6 bytes) of memory and could therefore be handled by a modern smartphone. This compressed representation drastically speeds up the computation process. Our method is up to 10,000 times faster than DLX at finding all solutions to exact cover problems. Moreover, our algorithm is the first one capable of finding all the solutions to a tetromino tiling problem with a 12×12 board size. We confirmed that there are 13,664,822,582,333,502,156,627,512 solutions to the problem.

3. Application to real-world problems

Since our method represents the set of found solutions as a ZDD, we quickly identify solutions that satisfy several conditions. With this feature, our method can find good solutions to real-world problems. For example, the problem of designing the layout of an apartment is known to be an exact cover problem [3]. By using our algorithm to construct a ZDD that represents the set of all possible floor plans, and then interactively adding conditions that match the buyer's requirements, we can efficiently find satisfactory arrangements. A demonstration system that uses our algorithm to find acceptable apartment layouts is shown in **Fig. 4**. Because our algorithm can find all possible room arrangements for an apartment, we can browse them as a list. Moreover, we can add conditions on possible floor plans and efficiently identify floor plans that satisfy the additional conditions.

4. Conclusion

Our new algorithm for solving exact cover problems is fast and can efficiently store all the solutions. These features enable the interactive discovery of desirable solutions by setting conditions in practical situations, which is especially beneficial for problems such as finding a desired floor plan. We are planning to improve the data structure in order to find the optimum solutions to practical exact cover problems.

(a) List all possible room layouts for a 3-bedroom apartment.

(b) Impose additional constraints for narrowing candidate solutions.

(c) List the solutions that satisfy constraints.

Fig. 4. Demonstration system for designing layouts of an apartment.

References

[1] D. E. Knuth, "Dancing Links," Millennial Perspectives in Computer Science, pp. 189–214, 2000.
[2] M. Nishino, N. Yasuda, S. Minato, and M. Nagata, "Dancing with Decision Diagrams: A Combined Approach to Exact Cover," Proc. of the 31st AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, Feb. 2017.
[3] A. Takizawa, Y. Miyata, and N. Katoh, "Enumeration of Floor-plans Based on a Zero-suppressed Binary Decision Diagram," International Journal of Architectural Computing, Vol. 13, No. 1, pp. 25–44, 2015.

Masaaki Nishino
Research Scientist, NTT Communication Science Laboratories.

He received a B.E., M.E., and Ph.D. in informatics from Kyoto University in 2006, 2008, and 2014. He joined NTT in 2008. His current research interests include data structures, natural language processing, and combinatorial optimization.

Norihito Yasuda
Senior Researcher, NTT Communication Science Laboratories.

He received a bachelor's degree in integrated human studies and a master's degree in human and environmental studies from Kyoto University in 1997, and 1999, and a D.Eng. in computational intelligence and system science from the Tokyo Institute of Technology in 2011. He joined NTT in 1999. He also worked as a research associate professor with the Graduate School of Information Science and Technology, Hokkaido University, in 2015. His current research interests include discrete algorithms and natural language processing.

Shin-ichi Minato
Professor, Graduate School of Information Science and Technology, Hokkaido University.

He received a B.E., M.E., and D.E. in information science from Kyoto University in 1988, 1990, and 1995. He worked in the NTT laboratories from 1990 until 2004. He was a visiting scholar in the Computer Science Department at Stanford University, USA, in 1997. He joined Hokkaido University as an associate professor in 2004, and has been a professor since October 2010. He has also worked as a visiting professor at the National Institute of Informatics since 2015. His research interests include efficient representations and manipulation algorithms for large-scale discrete structures such as Boolean functions, sets of combinations, sequences, and permutations. He was a research director of the JST ERATO MINATO Discrete Structure Manipulation System Project from 2009 to 2016 and is now leading a Grant-in-Aid for Scientific Research (KAKENHI) project of the Japan Society for the Promotion of Science (JSPS) until 2020. He is a senior member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Information Processing Society of Japan (IPSJ), and a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Japanese Society for Artificial Intelligence (JSAI).

Masaaki Nagata
Senior Distinguished Researcher, Group Leader, NTT Communication Science Laboratories.

He received a B.E., M.E., and Ph.D. in information science from Kyoto University in 1985, 1987, and 1999. He joined NTT in 1987. His research interests include morphological analysis, named entity recognition, parsing, and machine translation. He is a member of IEICE, IPSJ, JSAI, the Association for Natural Language Processing, and the Association for Computational Linguistics.

Memory-efficient Word Embedding Vectors

Jun Suzuki and Masaaki Nagata

Abstract

Word embedding is a technique for identifying the semantic relationships between words by computer. Word embedding vectors enable computers to provide a *guess* similar to the intuition or common sense of human beings. This article introduces a method for reducing the required memory consumption of this important fundamental operation of word embedding vectors while maintaining the ability to calculate semantic relationships, which is an important property when this technique is applied to real world systems.

Keywords: natural language processing, deep learning, word embeddings

1. Calculating the meaning of words

Many challenges still need to be overcome to reach the stage where computers can accurately understand and manipulate natural language at the same level as human beings. In particular, it is a difficult problem for computers to correctly understand semantic relationships between words. However, a method called *word embedding*—a technique enabling computers to understand the semantic relationships between words—is attracting a lot of attention from researchers and engineers in the natural language processing field (**Fig. 1**).

This technique was originally developed in the 1980s [1], but there has been a revival with the recent development of deep learning methods. More precisely, the method proposed by Mikolov et al. [2] has empirically proven that word embedding vectors can be trained within a feasible run-time even if the size of training data is very large, for example, web-scale data. Thus, the method can capture very accurate semantic relationships between words with the help of large-scale text data since such large-scale data should implicitly contain information equivalent to the *common sense* of human beings.

As a simple example for explaining the usefulness of word embedding vectors, computers can estimate the meanings of words by the vector calculations among the word embedding vectors. Suppose we ask someone "Which word has the most appropriate relation to the word 'Germany,' if the word is based on the same relation existing between 'France' and 'wine'?" Many people would answer "beer," for instance. Of course there is no unique correct answer for this question, and some people might say that "beer" is not a correct answer in his/her view. However, many people feel that "beer" is an acceptable answer.

These days, computers are becoming capable of developing such common sense or knowledge of human beings with the help of word embedding vectors. The most important point here is that this type of intuitive guess—similar to that done by human beings—is now manageable for computers.

Traditionally, the approach used by computers to identify the semantics of words involved the use of hand-made semantic dictionaries. The essential difference from such traditional dictionary-based methods with the word embedding approach is coverage and whether the method involves an automatic or hand-made construction. It is easily imaginable that the traditional dictionary-based methods can solve semantic problems with high accuracy if the dictionary has information on the given problems; if not, dictionary-based methods are not effective for solving such problems. Moreover, a large cost may be required to keep updating the dictionary to improve

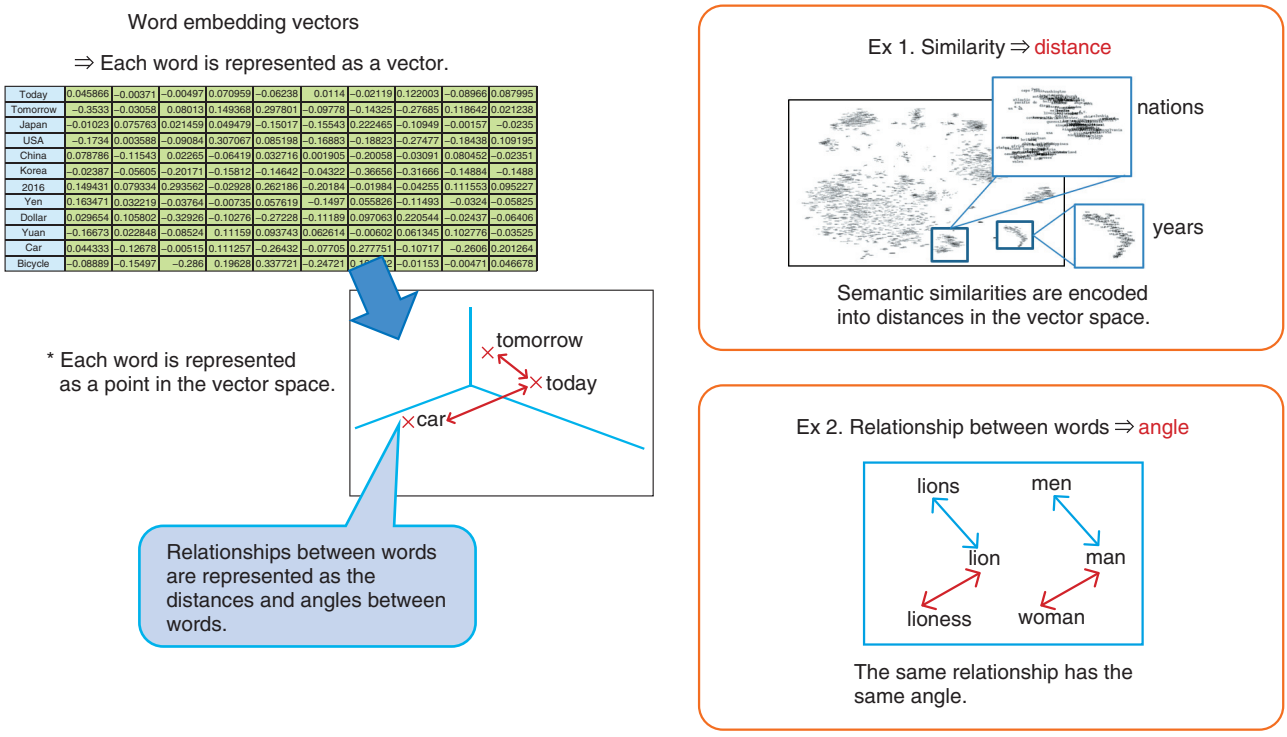


Fig. 1. Word embedding vectors.

the coverage.

For example, a computer might not be able to solve the above example of *beer* using a dictionary-based method since the importance of such common sense is relatively low, and thus, it might not be included in the dictionary. In contrast, word embedding vectors can be automatically generated from a large amount of text data, and no human cost is required for maintenance. Moreover, conceptually, information on all the words appearing in the training data can be stored in the embedding vectors. In fact, word embedding can easily handle millions of words. This fact implies that the word embedding method can handle a much larger number of words than dictionary-based methods (**Fig. 2**).

2. Usefulness of word embedding vectors for computers

Word embedding vectors can be utilized in many natural language processing applications such as machine translation, question answering, information retrieval, and document summarization. However, we sometimes encounter several inconvenient points when trying to apply word embedding vectors to real systems. For example, there are a lot of random factors when word embedding vectors are constructed using conventional methods. Therefore, the resultant word embedding vectors lack reproducibility, meaning they always differ from each other when many trials are conducted.

In another example, we have to completely retrain embedding vectors in situations where we need embedding vectors with distinct numbers of dimensions, since the dimensions of word embedding vectors can be pre-defined before starting the training, and different applications often prefer their own numbers of dimensions. This is an example indicating that an advanced technology cannot always be easily applied to real world systems. To overcome this inconvenience of low usability, we have developed several methods that have high usability [3–5]. In the remainder of this article, we explain one of our methods for significantly reducing the memory requirements of word embedding vectors [5].

3. Method for reducing memory requirements

We first explain the usefulness of reducing memory requirements. Suppose we are building a dialogue

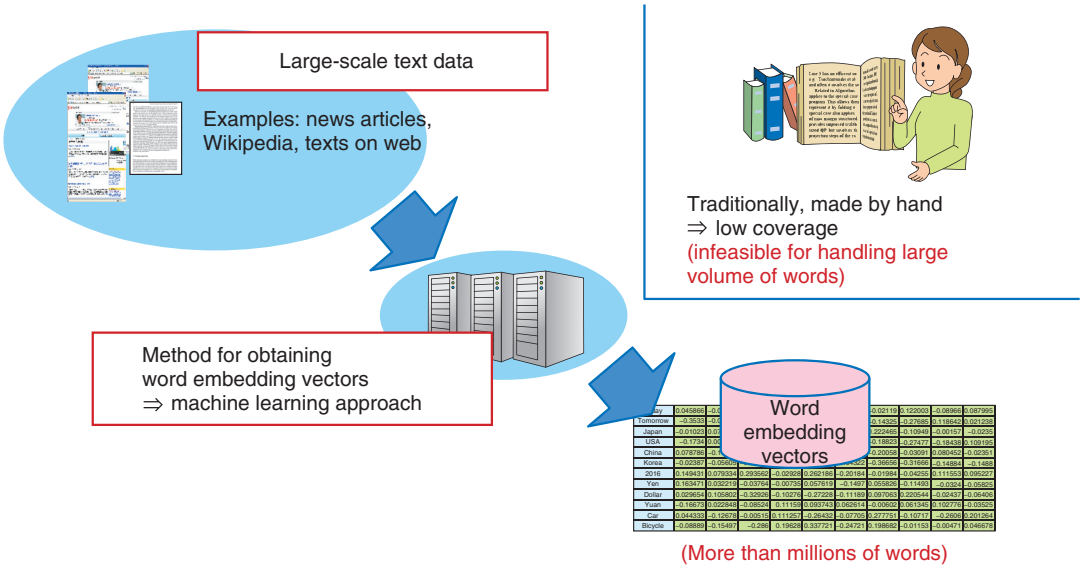


Fig. 2. Method for obtaining word embedding vectors.

system in robots used mainly for communicating with users. In this case, we aim to put as many words into the system as possible since it is nearly impossible for the system to appropriately process unknown words. However, there is a trade-off between performance and memory requirements in general. Namely, the amount of required memory storage to store all of the word embedding vectors becomes a large problem when we add a large number of words into the systems.

For example, let us consider the case when utilizing three million words and a 300-dimensional vector is assigned to each word. Here, we assume that we need a 4-byte memory to represent a single precision floating point number. Then the memory requirement for representing overall word embedding vectors becomes 3,000,000 (words) x 300 (dimensions) x 4 (bytes) = 3,600,000,000 (bytes). This means that it requires 3.4 GB of memory. We emphasize that we need 3.4 GB of memory only for a *single* module, not for an entire system. This is unacceptably large in general.

Here, we assume that memory requirements are one-hundredth, that is, 34 MB, of the above amount. Then the total cost of memory storage integrated into robots can be significantly reduced. This actual cost reduction is essentially the most important factor in a real world product. In addition, we can easily integrate word embedding vectors into applications on mobile devices. Less memory usage also leads to lower power consumption even though the memory storage in mobile devices may increase rapidly in the near future. Consequently, we can expect various positive effects for real world applications by merely developing a means of reducing the memory requirements.

4. Method

Our method for reducing the memory requirements consists of a combination of several machine learning techniques such as group regularization, dual decomposition, augmented Lagrangian methods, and clustering. We do not describe these techniques in detail here but rather briefly explain the essence of our method. As described previously, word embedding vectors are generated from large training data. More precisely, what the method is actually trying to do during the learning process is to find appropriate values in the embedding vectors assigned to each word by minimizing the given objective function. The basic idea of our method is as follows: Suppose we observe that a certain value sequence pattern, for example, (0.3, –0.2, 0.1, 0.5), appears many times in the obtained embedding vectors. In this situation, we can discard these patterns with no information loss by preserving a single value sequence pattern among them and adding information consisting of *appearance of the same value sequence pattern* to the locations where all the same value sequence patterns

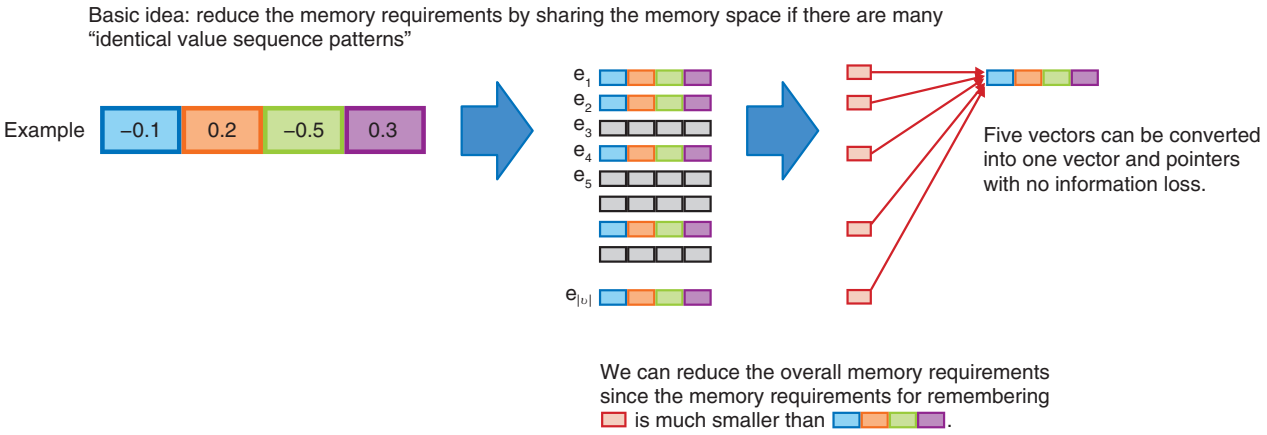


Fig. 3. Idea for reducing the memory requirements.

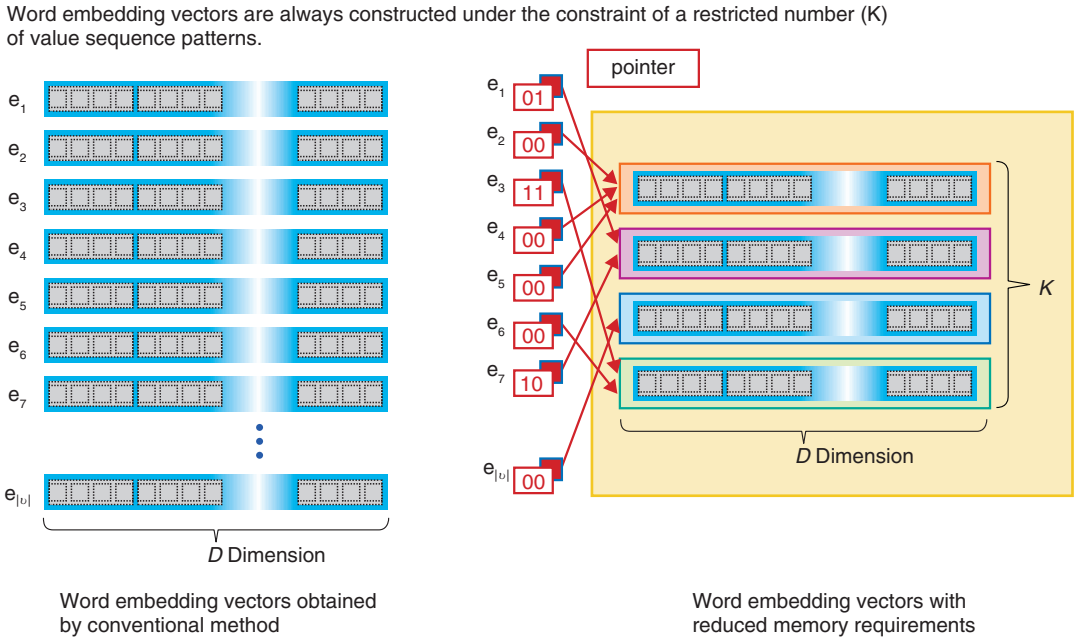


Fig. 4. Proposed method.

appear. Then, we can reduce the overall memory requirements if the memory requirement of remembering where the value sequence patterns appear is smaller than remembering the original value sequence patterns (**Fig. 3**).

By implementing this idea, we can continue to reduce the overall memory requirements if the number of distinct value sequence patterns in the word embedding vectors gets smaller and smaller. Unfortunately, however, none of the conventional methods automatically generate such convenient value sequence patterns. Therefore, we have built a method that can force the system to produce the word embedding vectors under the condition of *constructing word embedding vectors with pre-defined K distinct value sequence patterns* while maintaining the performance. Using this method, we can produce word embedding vectors within the memory desired by the users (**Fig. 4**).

5. Future direction

We are conducting research with the objective of having the most advanced research results become basic technologies that are used in real world systems including artificial intelligence related systems. It is possible to directly and indirectly support improvements of actual systems being used by further developing the basic technologies used in the systems. Thus, our final goal is to develop many basic technologies that offer high usability for computers and system developers, which we believe to be one of the most important characteristics of basic technologies.

References

[1] G. E. Hinton, "Learning Distributed Representations of Concepts," Proc. of the Eighth Annual Conference of the Cognitive Science Society, pp. 1–12, Amherst, MA, USA, Aug. 1986.

[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Proc. of the First International Conference on Learning Representations (ICLR 2013), Scottsdale, AZ, USA, May 2013.

[3] J. Suzuki and M. Nagata, "A Unified Learning Framework of Skip-grams and Global Vectors," Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference of the Asian Federation of Natural Language Processing (ACL-ICNLP2015), Beijing, China, July 2015.

[4] J. Suzuki and M. Nagata, "Right-truncatable Neural Word Embeddings," Proc. of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT2016), San Diego, CA, USA, June 2016.

[5] J. Suzuki and M. Nagata, "Learning Compact Neural Word Embeddings by Parameter Space Sharing," Proc. of the 25th International Joint Conference on Artificial Intelligence (IJCAI-16), New York, USA, July 2016.

Jun Suzuki

Senior Research Scientist, Linguistic Intelligence Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. in engineering from the Graduate School of Information Science, Nara Institute of Science and Technology in 2005. He joined NTT Communication Science Laboratories in 2001, where he is researching machine learning, natural language processing, and artificial intelligence areas.

Masaaki Nagata

Senior Distinguished Researcher, Group Leader, NTT Communication Science Laboratories.

He received a B.E., M.E., and Ph.D. in information science from Kyoto University in 1985, 1987, and 1999. He joined NTT in 1987. His research interests include morphological analysis, named entity recognition, parsing, and machine translation. He is a member of the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, the Japanese Society for Artificial Intelligence, the Association for Natural Language Processing, and the Association for Computational Linguistics.

Synthesizing Ghost-free Stereoscopic Images for Viewers without 3D Glasses

Taiki Fukiage, Takahiro Kawabe, and Shin'ya Nishida

Abstract

When a conventional stereoscopic display is viewed without three-dimensional (3D) glasses, image blurs, or *ghosts*, are visible due to the fusion of stereo image pairs. This artifact severely degrades 2D image quality, making it difficult to simultaneously present clear 2D and 3D content. To overcome this limitation, we recently proposed a method to synthesize ghost-free stereoscopic images. Our method gives binocular disparity to a 2D image and drives human binocular disparity detectors by the addition of a quadrature-phase pattern that induces spatial subband phase shifts. The disparity-inducer patterns added to the left and right images are identical except for the contrast polarity. Physical fusion of the two images cancels out the disparity-inducer components and makes only the original 2D pattern visible to viewers without glasses.

Keywords: stereoscopy, spatial phase shift, backward compatible

1. Introduction

In standard three-dimensional (3D) stereoscopic displays, a stereo pair (a pair of images each representing a view from each eye position) is simultaneously presented on the same screen. For viewers wearing 3D glasses, an image for the left eye and an image for the right eye are separately presented to the corresponding eyes, and viewers can enjoy 3D depth impressions due to binocular disparity (small differences between the image pair). For viewers who unfortunately watch the display without 3D glasses, however, the superposition of the left and right images produces uncomfortable image blurs and bleeding, which are referred to as *ghosts* in this article.

Stereoscopic ghosts, caused by the binocular disparity between the stereo image pair, seriously limit the utility of the standard 3D display. In particular, ghosts make the display unsuitable in situations where the same screen is viewed by a heterogeneous group of people, including those who are not supplied with stereo glasses or who have poor stereopsis (the perception of depth produced by the reception in the brain of visual stimuli from both eyes in combination).

2. Stereo image synthesis to achieve perfect backward compatibility

We recently proposed a method to synthesize completely ghost-free stereoscopic images [1]. Our technique gives binocular disparity to a 2D image by the addition of a quadrature-phase pattern that induces spatial subband phase shifts (**Fig. 1**). The patterns added to the left and right images, which we call disparity inducers, are identical except that they are opposite in contrast polarity. Physical fusion of the two images cancels out the disparity-inducer components and brings the image back to the original 2D pattern (**Fig. 2**). This is how binocular disparity is hidden from viewers without glasses but remains visible to viewers with them.

3. Technical details

Our method synthesizes a stereo image by linearly

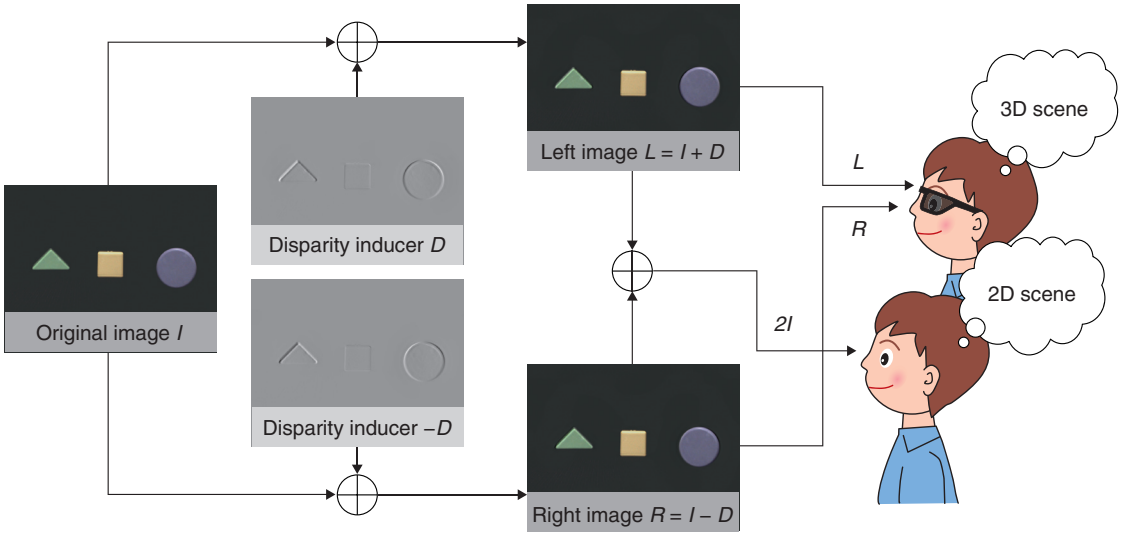


Fig. 1. Schematic of how our method cancels stereo ghosts for viewers without 3D glasses.

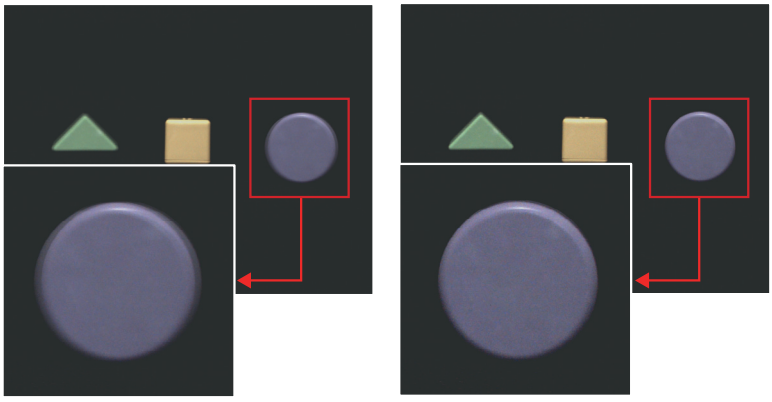


Fig. 2. Comparison of image quality when viewed without 3D glasses.

adding a disparity inducer. To realize this process, we manipulate binocular disparity based on a local phase shift. For the sake of simplicity, let us assume a vertical grating whose intensity profile is a sinusoidal wave to be an original image. We use a quadrature (half-cycle)-phase-shifted component of the original image as a disparity inducer (**Fig. 3**). When the quadrature-phase-shifted component is added to the original image, the resulting composite image is a quarter-cycle-phase-shifted wave from the original one.

By contrast, when the quadrature-phase-shifted component is subtracted from the original image, the phase of the original wave is shifted in the opposite direction by a quarter cycle. This phase difference between the resulting images is used as a disparity for producing a depth impression.

Importantly, we can control the size of the phase shift by modulating the amplitude of the quadrature-phase component. Therefore, we can reproduce the desired amount of disparity. The amount of phase shift required to reproduce a specific size of disparity depends on the spatial frequency and the orientation of the original grating. Fusion of the composite images cancels out the effects of the quadrature-phase component and makes the image profile the

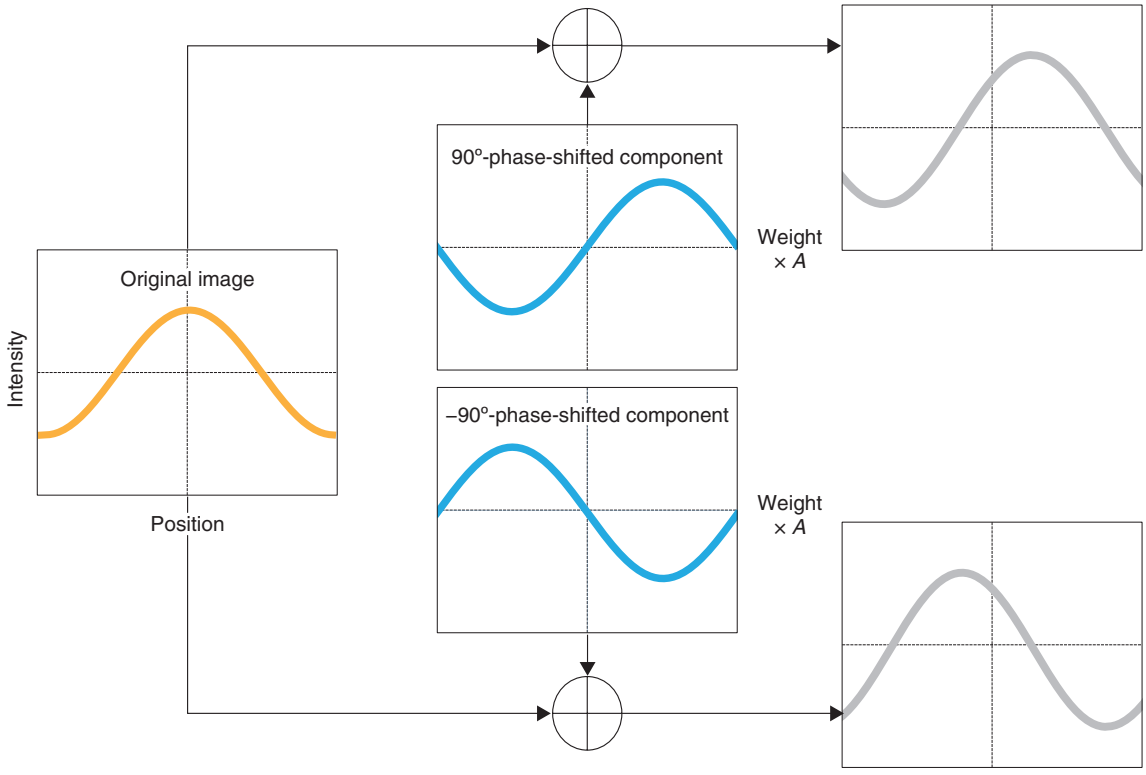


Fig. 3. Basic mechanism to produce a relative phase shift.

same as the original one.

4. Application to general images

In general, any arbitrary image can be represented as a combination of sinusoids by a 2D Fourier transform. Thus, the disparity manipulation described above can be easily applied to arbitrary images as long as the disparity for each sinusoidal component is spatially uniform. However, when we want to reproduce a structured map of binocular disparity (disparity map), we need spatially localized information about the spatial frequency and orientation of an input image. To analyze this local structure information, we use the steerable pyramid, which decomposes an image into subband images, each representing local responses tuned to different spatial frequencies and orientations [2].

The process to generate a disparity inducer is as follows: (1) The system first decomposes an input image into subband images using the steerable filters and obtains quadrature-phase-shifted versions of the subband images. (2) The quadrature-phase patterns are weighted so as to reproduce disparity sizes given by an input disparity map. The weight sizes are computed based on the spatial frequency and orientation of each of the quadrature-phase patterns. (3) Finally, the disparity inducer is obtained by reconstructing the weighted quadrature-phase subbands.

There is a huge amount of 3D content in the conventional stereo format. We can also transform that content into our ghost-free format. The most straightforward way to achieve this is to compute a disparity map from a given stereo pair. Then, one can generate a disparity inducer in the same way as described above using one of the stereo images as an original 2D image to which a disparity inducer is added.

However, we present here a more concise way, in which we compute disparity as a phase difference between a given stereo pair and calculate the weight for the quadrature-phase-shifted image directly from the phase difference. When we use the left image of an input stereo pair as an original 2D image to which the disparity is added, the process to generate a disparity inducer is as follows (**Fig. 4**): (1) The system first decomposes both of the input stereo images into a series of subband images using steerable filters. (2) Then we compute the phase difference between each

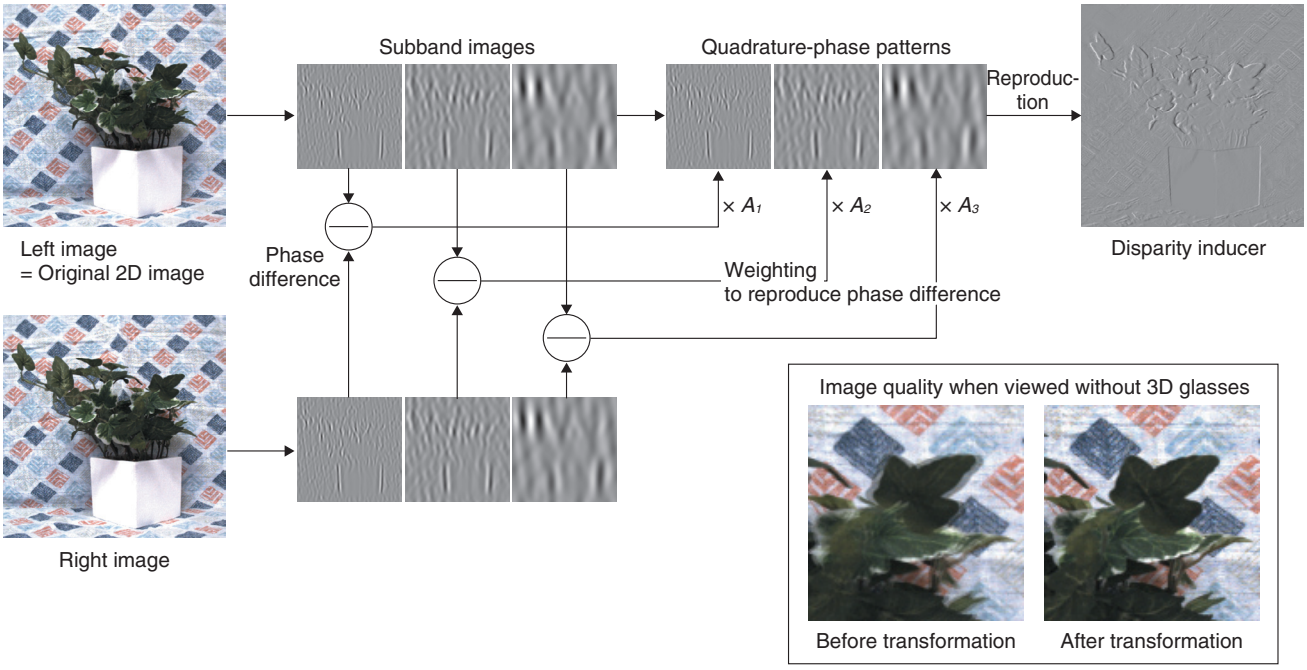


Fig. 4. Process overview to generate a disparity inducer from a standard stereo pair.

of the corresponding subband image pairs. (3) From each of the subband images of the left image, we extract the quadrature-phase component. (4) The quadrature-phase patterns are weighted so as to reproduce the phase differences obtained in (2). (5) The disparity inducer is obtained by reconstructing the weighted quadrature-phase subbands. After the disparity inducer is given, the left and right stereo images in our stereo format are respectively generated by the addition/subtraction of the disparity inducer to/from the input left image. By processing every frame in the same manner, we can transform not only static stereo pictures, but also stereo movie sequences into our ghost-free format.

The image processing described above is designed so as to effectively drive the binocular mechanisms in our brain. The human visual system analyzes binocular disparity using a bank of multi-scale, orientation-tuned bandpass mechanisms similar to the steerable pyramid. Then the visual system computes the binocular disparity as the phase difference between the corresponding bandpass responses. Therefore, our technique can properly drive the viewer's binocular vision by adding disparity as a spatial phase shift even though the synthesized stereo images are slightly different from those obtained in the real 3D world.

5. Comparison to previous approach

A previous solution to simultaneously present 3D content and ghost-free 2D content with a single display is to make a special image combination that cancels out, say, the right image component and leaves only the left image to viewers without 3D glasses [3, 4]. However, this method significantly sacrifices the contrast of the original image. In addition, it is incompatible with the standard stereo display system.

Another approach to attenuate stereo ghosts is to compress the magnitude of binocular disparity [5]. This approach is compatible with the standard stereo displays, but it cannot perfectly remove the stereo ghosts.

There is also a stereo-image-synthesis method that is similar to ours [6]. In this method, an edge-enhancing (derivative) filter is used to produce pseudo-disparity. However, this method cannot precisely control the size of disparities since it does not analyze the spatial frequency or orientation of the original image.

Our method can completely remove stereo ghosts while maintaining compatibility with standard stereo display hardware. Unlike the previous solution in which one of the stereo images is entirely canceled out, only the disparity inducer embedded in a stereo

image pair is canceled out in our case. Thus, the reduction in the contrast of the original image is moderate or even unnecessary. Moreover, in our method, the size of the phase shift is controlled based on the spatial frequency and orientation of the original image. Therefore, we can easily reproduce natural depth structures that are indistinguishable to the physically correct stereo images.

6. Limitation and future work

In exchange for the ability to present perfect ghost-free 2D images to viewers without glasses, our method has a slightly limited ability to present 3D images compared with the conventional position-shift method. Specifically, there is an upper limit in the disparity magnitude given to the image. Adding a large disparity could affect the dynamic range and induce perceptual artifacts. Thus, when transforming standard stereo images to our ghost-free format, one may have to combine it with a depth compression technique [5].

On the other hand, our method can provide 3D information in a slightly different way from the conventional methods because we can hide disparity information in 2D images. For example, one can use 3D glasses like a loupe to inspect complicated structures of an image on a personal computer monitor while looking at it without glasses otherwise.

In addition, since our technique is based on the linear addition of disparity inducers, when combined with projection mapping, it can add to real objects depth structures that are visible only to viewers with glasses. The projection system can be used, for example, in an art museum to present additional depth impressions onto existing paintings without causing any conflicts with observers who want to appreciate their original appearances.

Finally, our method provides a simple way to linearly decompose a stereo image pair into a single image and a depth pattern (disparity inducer). If the disparity inducer is given, it is easy to transform any 2D content into 3D content. It remains for investigation whether this decomposition contributes to other applications such as effective image data compression.

References

[1] T. Fukiage, T. Kawabe, and S. Nishida, “Hiding of Phase-based Stereo Disparity for Ghost-free Viewing Without Glasses,” ACM Trans. Graph. (Proc. of SIGGRAPH 2017, Los Angeles, CA, USA), Vol. 36, No. 4, pp. 147:1–17, 2017.

[2] E. P. Simoncelli and W. T. Freeman, “The Steerable Pyramid: A Flexible Architecture for Multi-scale Derivative Computation,” Proc. of the 2nd IEEE International Conference on Image Processing, Vol. 3, pp. 444–447, Washington, D.C., USA, Oct. 1995.

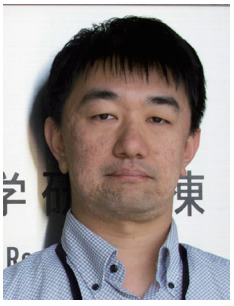
[3] W. Fujimura, Y. Koide, R. Songer, T. Hayakawa, A. Shirai, and K. Yanaka, “2x3D: Real Time Shader for Simultaneous 2D/3D Hybrid Theater,” Proc. of SIGGRAPH Asia 2012 Emerging Technologies, pp. 1:1–2, Singapore, Nov./Dec. 2012.

[4] S. Scher, J. Liu, R. Vaish, P. Gunawardane, and J. Davis, “3D+2DTV: 3D Displays with No Ghosting for Viewers Without Glasses,” ACM Trans. Graph., Vol. 32, No. 3, pp. 21:1–10, 2013.

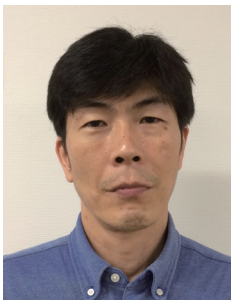
[5] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, “Apparent Stereo: The Cornsweet Illusion Can Enhance Perceived Depth,” Proc. of SPIE, Vol. 8291, Human Vision and Electronic Imaging XVII, IS&T/SPIE’s Symposium on Electronic Imaging, pp. 1–12, Burlingame, CA, USA, 2012.

[6] S. Kobayashi, “Image Processing Apparatus, Image Processing Method, and Program,” US Patent, US20100033554A1, 2010.

Taiki Fukiage
Research Associate, Sensory Representation Group, Human Information Science Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in interdisciplinary information studies from the University of Tokyo in 2015. He joined NTT Communication Science Laboratories in 2015, where he studies media technologies based on scientific knowledge about visual perception. He is a member of the Vision Sciences Society and the Vision Society of Japan.



Takahiro Kawabe
Senior Research Scientist, Sensory Representation Group, Human Information Science Laboratories, NTT Communication Science Laboratories.
He received a Doctor of Psychology from Kyushu University, Fukuoka, in 2005. In 2011, he joined NTT Communication Science Laboratories, where he studies human material recognition and cross-modal perception. He received the 2013 JPA Award for International Contributions to Psychology: Award for Distinguished Early and Middle Career Contributions from the Japanese Psychological Association. He is a member of the Vision Sciences Society and the Vision Society of Japan.



Shin'ya Nishida
Senior Distinguished Scientist, Group Leader of Sensory Representation Group, Human Information Science Laboratories, NTT Communication Science Laboratories.
He received a B.S., M.S., and Ph.D. in psychology from Kyoto University in 1985, 1987, and 1996. After a two-year stay at ATR Auditory and Visual Perception Laboratories as a research associate, he joined NTT in 1992. He is an expert in psychophysical research on human visual processing, in particular motion perception, cross-attribute/modality integration, time perception, and material perception. He is the current president of the Vision Society of Japan and an editorial board member of the Journal of Vision and Vision Research.

Personalizing Your Speech Interface with Context Adaptive Deep Neural Networks

Marc Delcroix, Keisuke Kinoshita, Atsunori Ogawa, Shigeki Karita, Takuya Higuchi, and Tomohiro Nakatani

Abstract

This article introduces our recent progress in speaker adaptation of neural network based acoustic models for automatic speech recognition. Deep neural networks have greatly improved the performance of speech recognition systems, enabling the recent widespread use of speech interfaces. However, recognition performance still greatly varies from one speaker to another. To address this issue, we are pursuing research on novel deep neural network architectures that enable rapid adaptation of network parameters to the acoustic context, for example, the speaker voice characteristics. The proposed network architecture is general and can potentially be used to solve other problems requiring adaptation of neural network parameters to some context or domain.

Keywords: deep learning, automatic speech recognition, speaker adaptation

1. Introduction

Automatic speech recognition (ASR) is being used more and more in our everyday life. For example, it is now common to speak to our smartphones to ask for the weather forecast or the nearest restaurant. Communication agents such as home assistants and robots are also starting to enter our living rooms, suggesting that speech may become a common modality for accessing information in the near future.

The rapid expansion of ASR based products has been made possible by the significant recognition performance gains achieved through the recent introduction of deep neural networks (DNNs) [1]. However, simply using DNNs does not solve all the issues. Speech recognition performance can still greatly vary depending on the acoustic context such as the speaker voice characteristics or the noise environment.

In this article, we describe our approach to tackle this problem by making the ASR system adaptive to the acoustic context. To achieve this, we have developed a novel DNN architecture that we call context adaptive DNN (CADNN) [2]. A CADNN is a neural network whose parameters can change depending on the external context information such as speaker or noise characteristics. This enables us to rapidly generate an ASR system that is optimal for recognizing speech from a desired speaker, opening the way to better ASR performance.

In the remainder of this article, we briefly review how current ASR systems work, focusing on the acoustic modeling part. We then describe in more detail the proposed CADNN and a speaker adaptation experiment we conducted to confirm its potential. We conclude this article by discussing some outlooks on potential extensions of CADNNs to achieve online speaker adaptation and applications to other research areas.

2. Deep learning based acoustic modeling

A speech recognition system is composed of several components, as illustrated in **Fig. 1**. First, there is a feature extraction module, which extracts speech

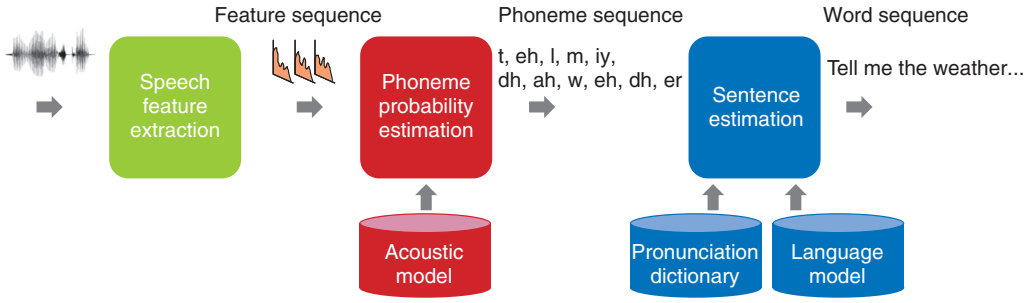


Fig. 1. Speech recognition system.

features from each short time frame of about 30 ms of a speech signal. Then, the acoustic model computes the probability that a speech feature corresponds to a given phoneme. Finally, the decoder finds the best word sequence given the input sequence of features by taking into account the phoneme probabilities obtained from the acoustic model, a pronunciation dictionary that maps the phoneme sequences to words, and scores obtained from the language model that outputs the probability of word sequences. In the remainder of this article, we focus our discussion on the acoustic model, and in particular on speaker adaptation.

Recently developed acoustic models use DNNs to map speech features to phoneme probabilities. An example of such an acoustic model is shown in **Fig. 2**. A DNN consists of several hidden layers that perform a nonlinear transformation of their input. With these stacked hidden layers, a DNN can model a complex mapping between its input features and its outputs. In the context of acoustic modeling, the inputs are speech features and the outputs are phoneme probabilities. Training such a DNN requires a large amount of speech data, from a few dozen hours to thousands of hours, depending on the task. The training data must also include the actual spoken phoneme sequences that can be derived from manual transcriptions of the utterances. With such training data, the acoustic model training follows the standard procedure for training DNNs such as error backpropagation with stochastic gradient descent.

To ensure that the acoustic model can well recognize speech in a variety of acoustic contexts such as for different speakers, the training data must contain speech from a large variety of speakers. Using such diverse training data enables us to obtain a good model on average. However, the DNN may not be optimal for a given speaker seen during the deployment of the recognition system because of the speaker's specific speaking style, which may result in poorer ASR performance for that particular speaker.

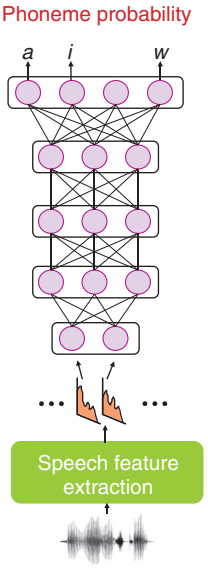


Fig. 2. DNN based acoustic model.

Solving this issue requires us to adapt the acoustic model to the desired speaker. However, adapting the acoustic model is challenging because it is often difficult to obtain the large amount of speech data with transcription that would be needed to train an acoustic model for the desired context. Specifically, it is impractical to require several hours of speech from each user to create a personalized acoustic model. In many applications, acoustic model adaptation should thus be *rapid*, that is, requiring a small amount of speech data such as a few seconds, and *unsupervised*, meaning it does not require transcribed data.

3. CADNN

Extensive research has been done to find approaches for adapting an acoustic model to speakers. A recent promising attempt consists of informing the DNN about the speaker by adding to its input an auxiliary feature describing the speaker characteristics. Such approaches have interesting properties because the speaker feature can be computed with only a few seconds of speech data, and they do not require transcriptions. However, simply adding an auxiliary feature to the input of a DNN has only a limited effect, as it can only partially adapt the DNN parameters. In this article, we describe an alternative way to exploit auxiliary information through a CADNN.

The idea behind CADNN is that a network trained for a given context should be optimal to recognize speech in that acoustic context. For example, we could build different networks to recognize speech from female and male speakers. Adaptation could then be realized simply by selecting the network corresponding to the target acoustic context. Such a naïve approach raises two issues. First, only part of the training data can be used for training each of the separate models. This would seem to be suboptimal because, for example, some speech characteristics are common to all speakers, and thus, better models could be trained when exploiting all the training data. Another issue is that it is unclear how to select the acoustic model in an optimal way.

The CADNN addresses these issues by making only part of the network dependent on the acoustic context. Moreover, we propose to select the model parameters using auxiliary features representing the acoustic context such as the speaker characteristics. A schematic diagram of a CADNN is shown in **Fig. 3** [3]. As illustrated in the figure, a CADNN has one hidden layer replaced by a context adaptive layer, that is, a layer that is split into several sublayers, each associated with a different acoustic context class.

For example, with two acoustic context classes, we could have a sublayer for male speakers and a sublayer for female speakers. The output of the hidden layer is obtained as a weighted sum of the output of each sublayer, with context weights derived from the auxiliary features. In our implementation, the context weights are computed from a small auxiliary network that has the auxiliary features as inputs. The outputs are the context weights that are optimal for recognizing speech for that acoustic context.

A CADNN has several interesting properties. The auxiliary network and the CADNN can be connected and trained jointly. This means that we can obtain context weights that are optimal for the acoustic context. Moreover, using such a joint training scheme, we do not need to explicitly define the acoustic context classes; they can be automatically learned from the training data during the training procedure. Finally, since except for the factorized layer, the rest of the network is shared among all the different acoustic context classes, all the training data can be used to train the parameters of the network.

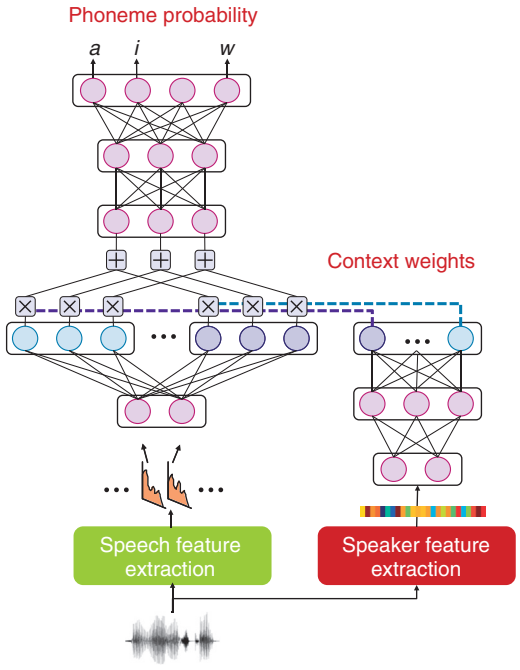


Fig. 3. Proposed CADNN for speaker adaptation.

4. Rapid speaker adaptation with CADNN

A CADNN can be used to achieve rapid speaker adaptation of acoustic models. The graph in **Fig. 4** shows the word error rate for recognition of English sentences read from the Wall Street Journal. Note that lower word error rates indicate better ASR performance. Our baseline system consists of a DNN with five hidden layers with ReLU (rectified linear unit) activations. The proposed CADNN uses a similar topology to that of the baseline DNN but has its second hidden layer replaced with a context adaptive layer, with four context classes. As auxiliary features, we use features representing speakers that are widely used for speaker recognition tasks. These auxiliary features were computed using a single utterance,

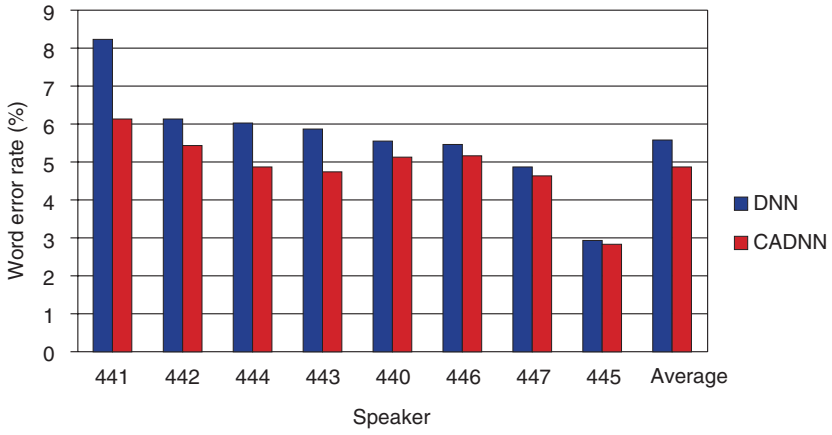


Fig. 4. Proposed CADNN for speaker adaptation.

which corresponds in this experiment to less than 10 s of speech data. Moreover, the speaker features can be obtained without transcriptions.

These results demonstrate that the proposed CADNN was able to significantly improve ASR performance, with a relative improvement of about 10% over the baseline. Since only a few seconds of speech data without transcriptions are sufficient to compute the auxiliary features, this experiment proves that CADNN can achieve rapid unsupervised speaker adaptation.

5. Outlook

The proposed CADNN appears promising for unsupervised rapid speaker adaptation of acoustic models. Potential further improvement could be achieved by developing better speaker representation for the auxiliary features [4]. Moreover, extension of the proposed scheme to online adaptation, where the adaptation process could start with even less data, is also a challenging research direction [5].

Finally, the proposed CADNN architecture is general and could be applied to other problems. For example, we are currently exploring the use of the same principle to extract a target speaker from a mixture of speakers [6]. We also believe that the proposed CADNN could be employed in other fields requiring context or domain adaptation of DNNs.

References

[1] Y. Kubo, A. Ogawa, T. Hori, and A. Nakamura, "Speech Recognition Based on Unified Model of Acoustic and Language Aspects of Speech," NTT Technical Review, Vol. 11, No. 12, 2013. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201312fa4.html

[2] M. Delcroix, K. Kinoshita, T. Hori, and T. Nakatani, "Context Adaptive Deep Neural Networks for Fast Acoustic Model Adaptation," Proc. of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4535–4539, South Brisbane, Australia, Apr. 2015.

[3] M. Delcroix, K. Kinoshita, C. Yu, A. Ogawa, T. Yoshioka, and T. Nakatani, "Context Adaptive Deep Neural Networks for Fast Acoustic Model Adaptation in Noisy Conditions," Proc. of ICASSP 2016, pp. 5270–5274, Shanghai, China, Mar. 2016.

[4] K. Veselý, S. Watanabe, K. Žmolíková, M. Karafiát, L. Burget, and J. Černocký, "Sequence Summarizing Neural Network for Speaker Adaptation," Proc. of ICASSP 2016, pp. 5315–5319, Shanghai, China, Mar. 2016.

[5] T. Ochiai, M. Delcroix, K. Kinoshita, A. Ogawa, T. Asami, S. Katagiri, and T. Nakatani, "Cumulative Moving Averaged Bottleneck Speaker Vectors for Online Speaker Adaptation of CNN-based Acoustic Models," Proc. of ICASSP 2017, New Orleans, USA, Mar. 2017.

[6] K. Žmolíková, M. Delcroix, K. Kinoshita, T. Higuchi, A. Ogawa, and T. Nakatani, "Speaker-aware Neural Network Based Beamformer for Speaker Extraction in Speech Mixtures," Interspeech 2017, Stockholm, Sweden, Aug. 2017.

Marc Delcroix
Senior Research Scientist, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

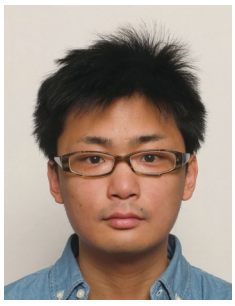
He received an M.Eng. from the Free University of Brussels, Brussels, Belgium, and the Ecole Centrale Paris, Paris, France, in 2003 and a Ph.D. from the Graduate School of Information Science and Technology, Hokkaido University, in 2007. He joined NTT Communication Science Laboratories as a research associate in 2007 and became a permanent research scientist in 2012. His research interests include robust multi-microphone speech recognition, acoustic model adaptation, integration of speech enhancement front-end and recognition back-end, speech enhancement, and speech dereverberation. He took an active part in the development of NTT robust speech recognition systems for the REVERB and CHiME 1 and 3 challenges, which all achieved best performance results in the tasks. He was one of the organizers of the REVERB challenge 2014 and of ASRU 2017 (Automatic Speech Recognition and Understanding Workshop). He is a visiting lecturer in the Faculty of Science and Engineering, Waseda University, Tokyo. He received the 2005 Young Researcher Award from the Kansai branch of the Acoustical Society of Japan (ASJ), the 2006 Student Paper Award from the Institute of Electrical and Electronics Engineers (IEEE) Kansai branch, the 2006 Sato Paper Award from ASJ, the 2015 IEEE ASRU Best Paper Award Honorable Mention, and the 2016 ASJ Awaya Young Researcher Award. He is a senior member of IEEE and a member of ASJ.

Keisuke Kinoshita
Senior Research Scientist, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received an M.Eng. and a Ph.D. from Sophia University, Tokyo, in 2003 and 2010. Since joining NTT in 2003, he has been engaged in research on speech and audio signal processing. His research interests include single- and multichannel speech enhancement and robust automatic speech recognition. He received Institute of Electronics, Information and Communication Engineers (IEICE) Paper Awards (2006), ASJ Technical Development Awards (2009), an ASJ Awaya Young Researcher Award (2009), Japan Audio Society Award (2010), and the Maejima Hisoka Award (2017). He is a member of IEEE, ASJ, and IEICE.

Atsunori Ogawa
Senior Research Scientist, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.E. and M.E. in information engineering and a Ph.D. in information science from Nagoya University, Aichi, in 1996, 1998, and 2008. He joined NTT in 1998. He is involved in researching speech recognition and speech enhancement. He received the ASJ Best Poster Presentation Awards in 2003 and 2006. He is a member of IEEE, the International Speech Communication Association (ISCA), IEICE, the Information Processing Society of Japan, and ASJ.



Shigeki Karita
Research Scientist, Media Information Laboratory, Signal Processing Research Group, NTT Communication Science Laboratories.

He received a B.Eng. and M.Eng. from Osaka University in 2014 and 2016. He joined NTT in 2016. His research interests include end-to-end speech recognition and speech translation. He received a Young Researcher Award from IEICE in 2014. He is a member of ASJ and ISCA.

Takuya Higuchi
Research Scientist, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.E. and M.E. from the University of Tokyo in 2013 and 2015. He has been with NTT since 2015, where he has been working on acoustic signal processing, array signal processing, blind source separation, and noise robust automatic speech recognition. He is a member of IEEE and ASJ.

Tomohiro Nakatani
Group Leader and Senior Distinguished Researcher, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.E., M.E., and Ph.D. from Kyoto University in 1989, 1991, and 2002. Since joining NTT as a researcher in 1991, he has been investigating speech enhancement technologies for intelligent human-machine interfaces. He spent a year at Georgia Institute of Technology, USA, as a visiting scholar in 2005. He has also been a visiting assistant professor in the Department of Media Science, Nagoya University, since 2008. He received the 2005 IEICE Best Paper Award, the 2009 ASJ Technical Development Award, the 2012 Japan Audio Society Award, the 2015 IEEE ASRU Best Paper Award Honorable Mention, and the 2017 Maejima Hisoka Award. He was a member of the IEEE SP Society Audio and Acoustics Technical Committee (AASP-TC) from 2009 to 2014 and served as the chair of the AASP-TC Review Subcommittee from 2013 to 2014. He is a member of the IEEE SP Society Speech and Language Processing Technical Committee (SL-TC). He served as an associate editor of the IEEE Transactions on Audio, Speech and Language Processing from 2008 to 2010 and was a chair of the IEEE Kansai Section Technical Program Committee from 2011 to 2012, a Technical Program co-Chair of IEEE WASPAA-2007, a Workshop co-Chair of the 2014 REVERB Challenge Workshop, and a General co-Chair of the IEEE Automatic Speech Recognition and Understanding Workshop. He is a member of IEEE, IEICE, and ASJ.

Optimization of Harvest Time in Microalgae Cultivation Using an Image Processing Algorithm for Color Restoration

Akihiro Kohno, Takeshi Komatsu, Soichi Oka, Hiroaki Fukuda, Hidehiko Yasui, and Hajime Makino

Abstract

Algae have been attracting attention as a next-generation alternative energy source to fossil fuels, but the high cost of cultivation remains an issue. This article introduces technology that reduces the cultivation cost through the use of an image-processing algorithm for color restoration to quickly determine the optimal time to harvest algae.

Keywords: image processing algorithm, microalgae, color restoration

1. Introduction

Global warming has become a serious environmental issue. The application of photosynthesis using microalgae has come under focus as way to reduce the amount of carbon dioxide in the atmosphere, which is deemed to be the cause of the greenhouse effect. Microalgae absorb carbon dioxide, convert it to oil, and store it internally. This oil features high added value with applications to biofuels, moisturizing oils, and health food (**Fig. 1**).

DENSO CORPORATION has begun cultivating a microalga called *Pseudochoricystis ellipsoidea*. Compared with terrestrial plants, microalgae have high growth potential and fast oil production on a par with light oils. These features make microalgae a good candidate for a next-generation, alternative energy source to fossil fuels. However, the current cultivation process is costly in comparison to the existing oil-producing process. In response to this issue, DENSO is conducting tests at a large-scale cultivation facility with the aim of reducing costs.

2. Key points in reducing cultivation costs

Algae have the property of stopping oil production after a certain growth period has been exceeded. Consequently, allowing cultivation to continue past this growth stop point results in unnecessary cultivation costs (**Fig. 2**). Harvesting at the growth stop point would therefore make for optimal cultivation.

We point out here that the total amount and type of

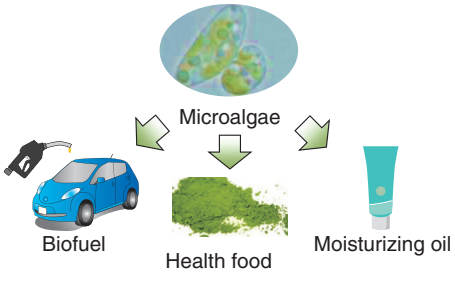


Fig. 1. Application areas for oil created by microalgae.

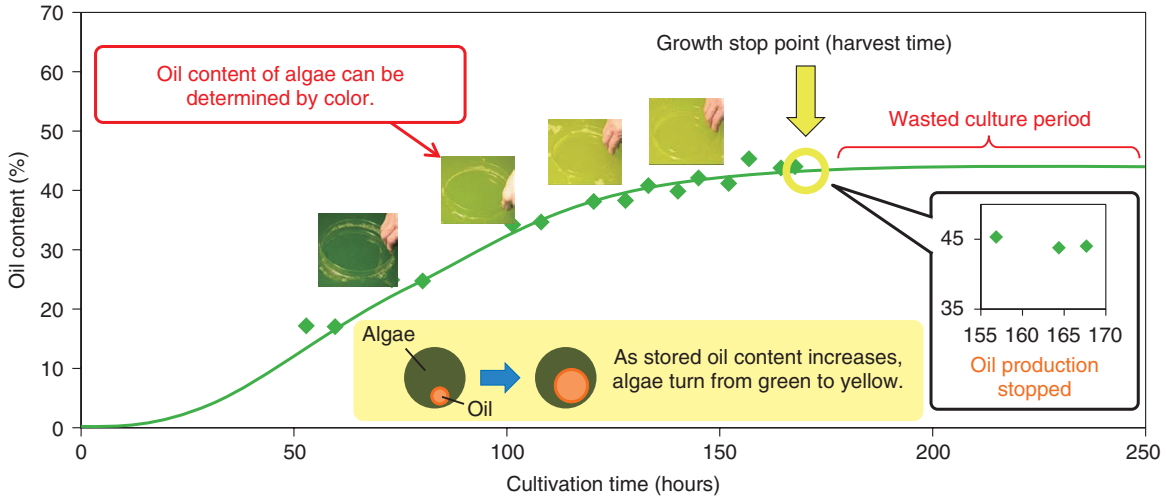


Fig. 2. Growth characteristics of algae.

oil stored by microalgae depend largely on the environment. As a result, a difference in the growing environment even when cultivating the same type of algae will cause the weight percentage of stored oil to differ, which means that this value cannot be used as a simple index for assessing the cultivation time and determining the growth stop point.

However, the color of the algae cultivation solution differs according to growth. In the early stage of cultivation, the number of cells is increasing, and the color green originating in chlorophyll, a photosynthesis pigment, is becoming darker. However, once the storage of oil begins and the constituent weight of that oil in the algae increases, the color of the algae itself begins to change from green to yellow.

Accordingly, if the stop point in this color change can be accurately identified, it should be possible to determine the time at which oil production stops.

3. Development of color determination technology at NTT

NTT has been developing color determination technology in relation to the concentration of chemical substances for some time and has applied this technology to environmental monitoring of ozone and other photochemical oxidants (**Fig. 3**). Furthermore, while studying methods for evaluating the growth of algae as a new application of this technology, we entered into a joint research project with DENSO.

The existing growth evaluation method involves sampling the algae cultivation solution and analyzing the components of the algae (**Table 1**). The entire process is done manually, which results in high labor costs and a time-consuming process that takes several days to complete from sampling to evaluation. The method introduced here, however, involves evaluating color using the RGB (red, green, and blue color model) information contained in an image of the algae. It enables simultaneous multipoint monitoring by capturing the algae over a broad area. In addition, connecting this method to a network enables remote operation and automatic measurements to be performed. Moreover, because image analysis can be performed in a short time with this technology, the time from determining the growth stop point to harvesting the algae can be shortened, substantially reducing the wasted cultivation period.

4. Technical issues in remote image analysis

Algae cultivation pools are installed outdoors, and therefore, reflections unrelated to components in the cultivation solution caused by foam on the solution surface and pool surroundings will affect the captured image. Additionally, because sunshine conditions can change from hour to hour, camera-shooting conditions cannot be fixed. These technical issues that prevent the correct determination of colors from the source image have an effect on remote image analysis (**Fig. 4**).

In response to the first issue, we suppressed reflections by installing a box coated with anti-reflective paint on the upper surface of the cultivation pool and

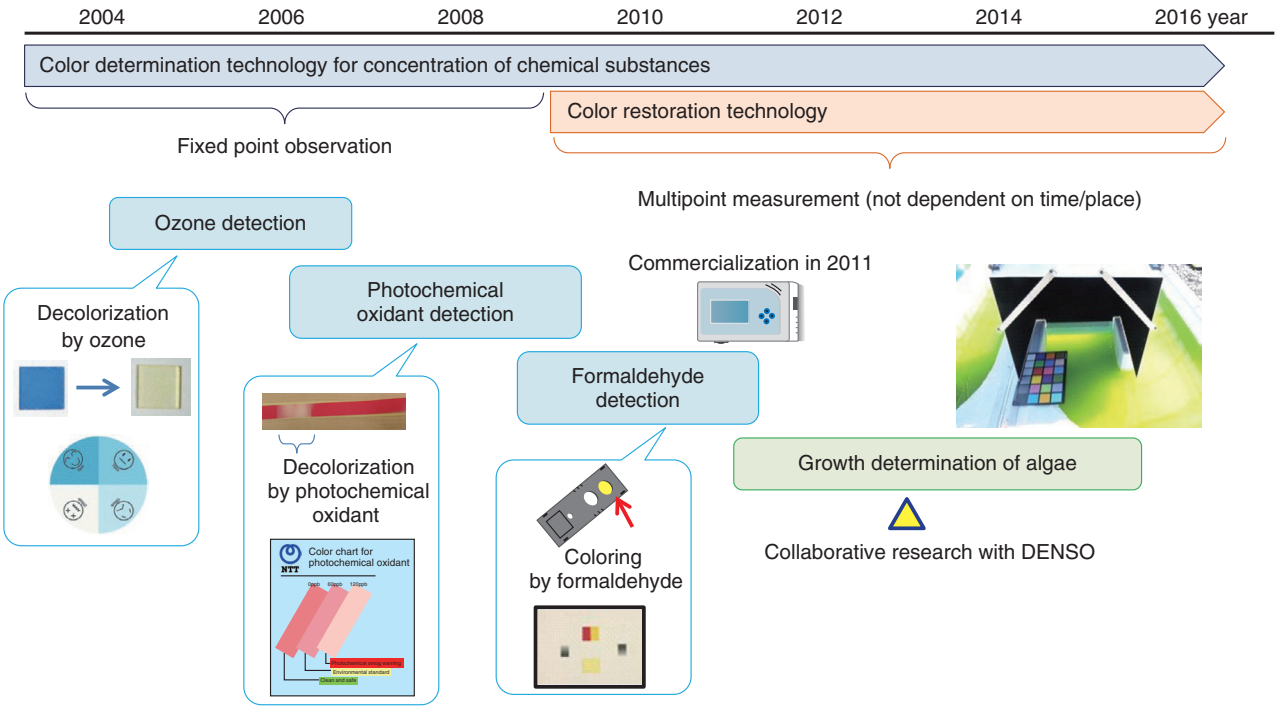


Fig. 3. Timeline of NTT research on color determination technology.

Table 1. Comparison of conventional and proposed methods.

	Conventional method	This method
Cultivation cost	High labor cost • Manual judgment • Long measurement time	Low labor cost • Automatic judgement • Short measurement time
Analysis method	Sampling, Filtration, Extraction, Centrifugation, Spectroscopic measurement	Image acquisition, Data transfer, Image analysis
Culture cycle	Not optimized Growth stop point, Harvest Oil content, Implantation, Time Wasted culture period	Optimized Growth stop point, Harvest Oil content, Implantation, Time

shooting the surface of the cultivation solution from only the inner side of the box.

We addressed the second issue by estimating the effects of sunshine at the time of image capture from RGB information and performing brightness correction and color-temperature correction against the image of the cultivation solution. The flow of these correction procedures is given below.

4.1 Step 1: Correction for sunshine brightness

Since colors cannot be accurately determined if the brightness of image information changes, we

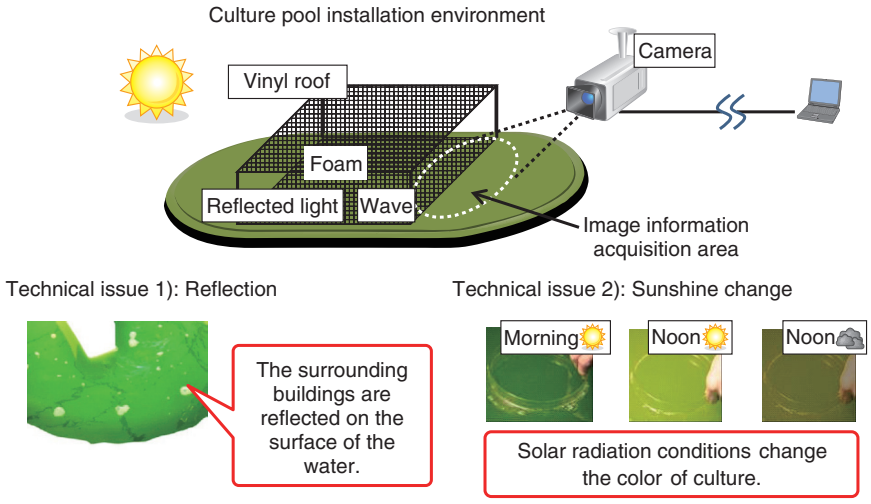


Fig. 4. Technical issues in remote image analysis.

standardized the brightness value using the following procedure. First, we obtained the R, G, and B gradation (0–255) for each of the pixels in the measurement area (**Eq. 1**).

$$\begin{aligned} &R_i\ (i = 0\text{–}255) \\ &G_i\ (i = 0\text{–}255) \\ &B_i\ (i = 0\text{–}255) \end{aligned} \tag{1}$$

Next, we created a histogram of frequency and gradation for the pixels in the measurement area and calculated the product of gradation and gradation frequency X for each color component (**Eq. 2**).

$$\begin{aligned} &R_i * X_{Ri} \\ &G_i * X_{Gi} \\ &B_i * X_{Bi} \end{aligned} \tag{2}$$

Now, by applying Eq. 2 to all pixels and taking the sum total, we calculated the brightness of the entire image, and by standardizing the brightness of each R, G, and B measurement area against the brightness of the entire image, we standardized the brightness of each R, G, and B component (**Eq. 3**).

$$\begin{aligned} R' &= \frac{\sum_{i=0}^{255} R_i X_{Ri}}{\sum_{i=0}^{255} R_i X_{Ri} + \sum_{i=0}^{255} G_i X_{Gi} + \sum_{i=0}^{255} B_i X_{Bi}} \\ G' &= \frac{\sum_{i=0}^{255} G_i X_{Gi}}{\sum_{i=0}^{255} R_i X_{Ri} + \sum_{i=0}^{255} G_i X_{Gi} + \sum_{i=0}^{255} B_i X_{Bi}} \\ B' &= \frac{\sum_{i=0}^{255} B_i X_{Bi}}{\sum_{i=0}^{255} R_i X_{Ri} + \sum_{i=0}^{255} G_i X_{Gi} + \sum_{i=0}^{255} B_i X_{Bi}} \end{aligned} \tag{3}$$

4.2 Step 2: Correction for sunshine color temperature

To evaluate the effect of a change in sunshine color temperature on the RGB information in the algae and to correct for this effect, we installed a color sample board (**Fig. 5**).

We then obtained the RGB information of the image of this color sample board beforehand (**Eq. 4**) as a reference for shooting conditions; here, R_n denotes the R information of the nth color sample.

$$\begin{pmatrix} R_1 & G_1 & B_1 \\ R_2 & G_2 & B_2 \\ R_3 & G_3 & B_3 \end{pmatrix} \tag{4}$$

Next, by obtaining the RGB information of the color sample board captured under the shooting conditions to be targeted for correction, we were able to denote the change in color with respect to the reference RGB information using 3×3 matrix T_{RGB} in **Eq. 5**, where R'_n denotes the R information of the nth color sample to be corrected.

$$\begin{pmatrix} R_1 & G_1 & B_1 \\ R_2 & G_2 & B_2 \\ R_3 & G_3 & B_3 \end{pmatrix} = T_{RGB} \begin{pmatrix} R'_1 & G'_1 & B'_1 \\ R'_2 & G'_2 & B'_2 \\ R'_3 & G'_3 & B'_3 \end{pmatrix} \tag{5}$$

Then, by applying the inverse matrix of this color-change matrix T_{RGB} to the RGB information of the cultivation solution targeted for correction (**Eq. 6**), we performed color restoration with respect to reference sunshine conditions. Here, R'_{algae} is R information of the cultivation solution before correction, and R''_{algae} is R information of the cultivation solution

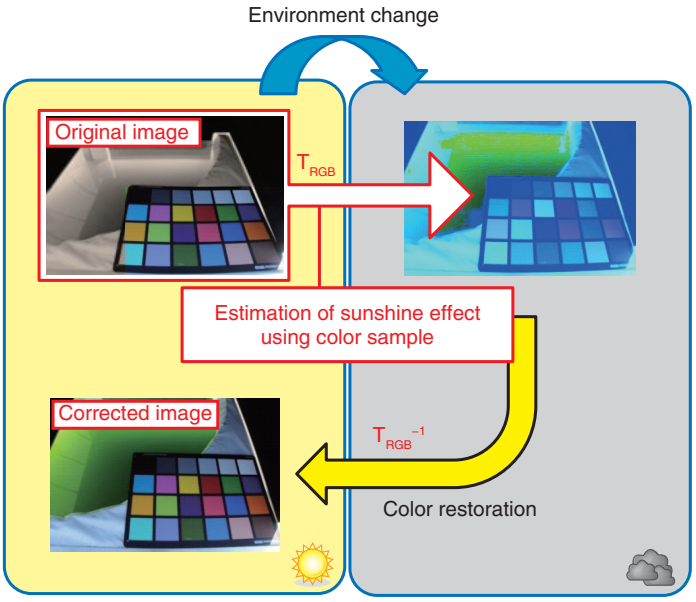


Fig. 5. RGB information correction using a color sample board.

after correction.

$$\begin{pmatrix} R''_{algae} \\ G''_{algae} \\ B''_{algae} \end{pmatrix} = T_{RGB}^{-1} \begin{pmatrix} R'_{algae} \\ G'_{algae} \\ B'_{algae} \end{pmatrix} \quad (6)$$

4.3 Step 3: Determination of harvest time

The color of the cultivation solution changes from green to yellow as the oil content percentage of the algae increases. In particular, we found that this increase in oil has a strong relationship with the percentage of R in the color of the solution. We therefore defined the percentage of R in the standardized RGB values using **Eq. 7** and took this value of r to be an index for determining the harvest time.

$$r = \frac{R'}{R' + G' + B'} \quad (7)$$

5. Experimental results and discussion

We collected some of the cultivation solution in an outdoor cultivation pool to compare the results of correcting for lighting conditions using the proposed method. The results shown in **Fig. 6** compare RGB values (Δ) under fixed lighting (indoors), RGB values (×) under sunshine (outdoors), and RGB values (O) of an image captured under sunshine and corrected using the proposed method.

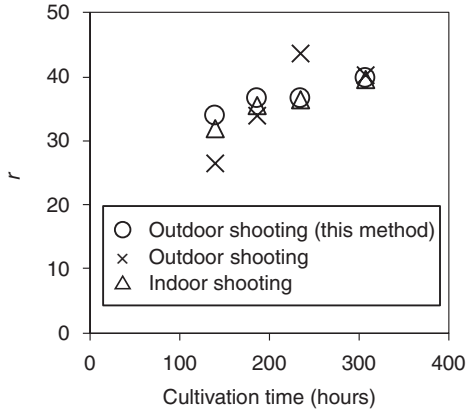
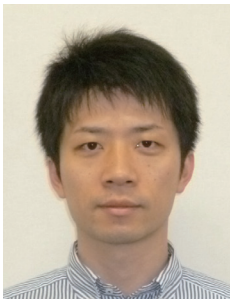


Fig. 6. Results of correction applying the proposed method.

An examination of these results indicates that the r value under sunshine and that under fixed lighting show the most divergence, but that the r value after correction using the proposed method is nearly the same as that under indoor shooting conditions. These results demonstrate that the proposed method can restore an image captured outdoors to practically the same image captured in a disturbance-free indoor environment.

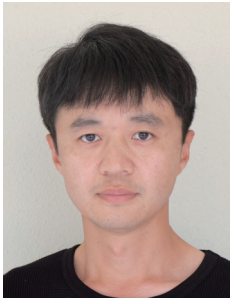
6. Conclusion

In this article, we introduced a method for determining in a short time the optimal harvest time for algae using a color restoration algorithm and obtained results that contribute to a reduction in cultivation costs. This method can be applied to any phenomenon for which a change in color occurs due to a change in constituent components. Going forward, we plan to expand the application domain of this method.



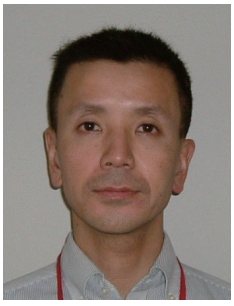
Akihiro Kohno
Research Engineer, Social Device Technology Laboratory, NTT Device Technology Laboratories.
He received a B.E. and M.E. in applied chemistry from the University of Tokyo in 2008 and 2010. He joined NTT Energy and Environment Systems Laboratories in 2010 and studied image restoration algorithms for application to the growth of microalgae. He is currently studying optical computing systems. Due to an organizational change in April 1, 2015, he is now with NTT Device Technology Laboratories. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Japan Society of Applied Physics (JSAP).

Takeshi Komatsu
Senior Research Engineer, Supervisor, Group Leader, NTT Device Technology Laboratories.
He received a B.S. and M.S. in physics from Yokohama City University, Kanagawa, in 1996 and 1998, and a Ph.D. in engineering from Nagaoka University of Technology, Niigata, in 2007. He joined NTT in 1998 and conducted research on fuel cells and in particular, solid-oxide fuel cells, up to 2010. He was with the NTT Environmental Protection Office from 2010 to 2013, after which he joined NTT Energy and Environment Systems Laboratories. He moved to the Social Device Technology Laboratory in 2014, where he researches materials technology for the production and storage of energy. He is a member of the Electrochemical Society of Japan.



Soichi Oka
Senior Research Engineer, Supervisor, Group Leader, NTT Device Technology Laboratories.
He received a B.E. in environmental information and an M.E. and Ph.D. in media and governance from Keio University, Kanagawa, in 1996, 1998, and 2001. During 2002–2003, he was engaged in research and development (R&D) of evanescent microwave imaging at Wright Patterson Air Force Research Laboratories and the University of Cincinnati in Ohio, USA. After joining NTT in 2004, he was involved in R&D of a millimeter-wave imaging system. His current research involves an inspection system for civil engineering structures.

Hiroaki Fukuda
Bio Research Department, Advanced Research and Innovation Division 4, DENSO CORPORATION.
He completed research on photosynthesis in coffee leaves at the Graduate School of Agricultural Science, Kobe University, in March 1992. He entered Nipponsanso Co. Ltd. in April 1992, where he was in charge of the development of stable isotope labeling reagents. He joined DENSO in August 1997 and worked on environmental biotechnology at Basic Research Laboratories. He is currently engaged in the R&D of biofuels using microalgae.



Hidehiko Yasui
Bio Research Section 1, Bio Research Department, Advanced Research and Innovation Division 4, DENSO CORPORATION.
He received a B.E., M.E., and Ph.D. in precision machinery engineering from The University of Tokyo in 2003. He was recognized for his paper "Development of an Electrostatic Levitation Linear Motor for Use in High Vacuum and Clean Environments." He joined DENSO in April 2003, where he worked on the development and design of high-voltage power control units. In April 2009, he was seconded to Toyota Motor Corporation to work on the development of production technology for high-voltage power control units. In July 2011, he returned to DENSO and began researching biofuels using microalgae.

Hajime Makino
Functional Testing Department Section, DENSO CORPORATION.
He graduated from Meinan Technical High School, Aichi, in March 1982. He joined DENSO in April 1982 and became the operator of the transfer line in the Meter Business Division. He transferred to Basic Research Laboratories in April 1997 and was in charge of silicon carbide (SiC) R&D. In January 2010, he began working on the development of biofuels using microalgae.

Report on First Meeting of ITU-T TSAG (Telecommunication Standardization Advisory Group) for the Study Period 2017 to 2020

Hideyuki Iwata

Abstract

The first meeting of the Telecommunication Standardization Advisory Group (TSAG) of the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) for the study period 2017 to 2020 was held at the ITU headquarters in Geneva, Switzerland, May 1–4, 2015, with some 120 delegates from 40 countries attending. The 11-person Japanese contingent comprised representatives of the Ministry of Internal Affairs and Communications, NTT, KDDI, Hitachi, Fujitsu, Mitsubishi Electric, NEC, OKI, and TTC (The Telecommunication Technology Committee). The new organization of TSAG is overviewed, and topics discussed during the meeting are described in this article.

Keywords: ITU-T TSAG, digital currency, blockchain

1. New organization of TSAG

It was decided that the Telecommunication Standardization Advisory Group (TSAG) will be made up of six Rapporteur Groups (RGs):

(1) Standardization Strategy (StdsStrat)
(2) Working Methods (WM)
(3) Work Programme and Study Group Structure (WP)
(4) Strengthening Cooperation/Collaboration (SC)
(5) Strategic and Operational Plan (SOP)
(6) Review of WTSA Resolutions[*1] (ResReview)

Yoichi Maeda, chief executive officer of TTC (The Telecommunication Technology Committee), was appointed rapporteur for the newly established RG-StdsStrat. Six associate rapporteurs from the telecommunications industry (Alibaba, Nokia, Ericsson, Tunisie Telecom, Cisco, and one person to be selected from Central and South America) were also appointed to this RG.

2. Establishment of Focus Groups on digital finance

The opening plenary meeting discussed establishment of two Focus Groups (FGs) on digital finance. The first was an FG on network infrastructure for digital fiat currency, which was jointly proposed by eCurrency Mint, several African countries, and AICTO (Arab Information and Communication Technologies Organization). The second was an FG on blockchains, which was proposed by Study Group (SG) 17 and South Korea. The US and UK representatives expressed concerns about the relations between the new FGs and the results of the FG on Digital Financial Services, which had completed its two-year study, and concerns about the need for the two new FGs and the relations between them.

*1 Resolution: The resulting document of the World Telecommunication Standardization Assembly (WTSA) giving guidance to the ITU-T structure, study programmes, and work methods.

Since a conclusion on this matter was not reached at the plenary meeting, the RG-StdsStrat took over the discussion. There was a dispute regarding the FG on digital fiat currency. The US in particular insisted on adding "network infrastructure" to the title of the FG because it believes that the FG should limit its study to network infrastructure. The eCurrency representative insisted that "fiat" be included. The title that was ultimately adopted was "Focus Group on Digital Currency including Digital Fiat Currency." Establishment of the FG was approved, and David Wen (eCurrency) was appointed as its chair. Regarding the FG on blockchains, it was decided that this group should address Digital Ledger Technology (DLT), which is a concept that is broader than blockchains, and that relations with other standards bodies (U4SSC[*2], ISO/TC307[*3], etc.) should be clarified. Ultimately, the "Focus Group on Application of Distributed Ledger Technology" was established, and David Watrin (Swisscom) was appointed as its chair.

3. Transfer of results of FG on Digital Financial Services

The FG on Digital Financial Services concluded its two-year study in December 2016, having produced 28 reports. The reports were sent to relevant SGs (SG2[*4], SG3[*5], SG12[*6], SG16[*7], and SG17[*8]). It was agreed that the focus of study on this matter will move to development of recommendations on service definitions, regulations, network requirements, security, and interoperability of digital financial services, within the ambit of ITU-T expertise.

4. Major discussions at each RG

Here, the main topics of discussion in each group and the conclusions reached are described.

4.1 RG-StdsStrat

It was agreed that the group's role is to advise TSAG and SGs on standardization strategies for the sector by identifying the main technological trends as well as the market, economic, and policy needs in ITU-T's fields of activity. Japan submitted a contribution introducing "Society 5.0."[*9] It proposed that each country's policy and the plans of other standards development organizations regarding standardization strategy be taken into consideration in developing ITU-T's standardization strategy.

4.2 RG-WM

This group reviewed resolutions and recommendations relating to working methods. On the basis of the discussion on working methods at WTSA-16, it was decided that problems identified in Resolution 1 and Recommendations A.1 and A.13 would be reviewed, and necessary revisions would be made. It was also agreed to study differences between what is specified in Recommendation A.1 and what is written in the rapporteur/editor manual.

4.3 RG-WP

The group approved introduction of new questions and revisions to existing questions. New and revised questions approved in this meeting are listed in **Table 1**. SG3's proposed new question, "economic and policy issues pertaining to QoS/QoE," encountered opposition from SG12. It was decided that SG3 and SG12 will jointly study the matter before the next TSAG meeting. Regarding which SG should lead the study on big data, it was decided not to identify any lead SG but rather, to encourage collaboration between SGs. Regarding obsolete work items (work items for which no progress had been made for 18 months), TSAG will send liaison statements to the respective SGs requesting each group to review such work items.

4.4 RG-SC

This group discussed collaboration with open source communities. Canada and the US indicated a need to define open source, to analyze collaboration

*2 U4SSC: United for Smart Sustainable Cities. This initiative was established by ITU and the United Nations Economic Commission for Europe.
*3 ISO/TC307: The Technical Committee on standardization of blockchain technologies and distributed ledger technologies of the International Organization for Standardization (ISO).
*4 SG2: Operational aspects of service provision and telecommunications management.
*5 SG3: Tariff and accounting principles and international telecommunication/ICT (information and communication technology) economic and policy issues.
*6 SG12: Performance, quality of service (QoS), and quality of experience (QoE).
*7 SG16: Multimedia encoding, systems, and applications.
*8 SG17: Security.
*9 Society 5.0: The Cabinet Office of the Government of Japan is promoting the creation of a super smart society (Society 5.0) under the theme "New Initiatives toward Japanese Industry of the Future and Social Transformation" as The 5th Science and Technology Basic Plan. Society 5.0 signifies how society has become progressively smarter through its transformation from a hunter-gatherer society to an agricultural society, industrial society, and information society, and the coming super smart society.

Table 1. New and revised questions approved at this meeting.

SG	Question number	Question title	Type
SG3	Q12/3	Tariffs, economic and policy issues pertaining to mobile financial services (MFS)	New
	Q13/3	Study of tariff, charging issues of settlements agreement of trans-multi-country terrestrial telecommunication cables	New
SG13	Q19/13	End-to-end cloud computing management, cloud security and big data governance	Revised
	Q20/13	IMT-2020: network requirements and financial architecture	Revised
	Q21/13	Network softwarization including software-defined networking, network slicing and orchestration	Revised
SG17	Q6/17	Security aspects of telecommunication services, networks and Internet of Things	Revised
	Q13/17	Security aspects for Intelligent Transport Systems	New
SG20	Q1/20	End to end connectivity, networks, interoperability, infrastructures and Big Data aspects related to IoT and SC&C	Revised and restructured
	Q2/20	Requirements, capabilities and use cases across verticals	
	Q3/20	Architectures, management, protocols and Quality of Service	
	Q4/20	e/Smart service, applications and supporting platforms	
	Q5/20	Research and emerging technologies, terminology and definitions	
	Q6/20	Security, privacy, trust, and identification for IoT and SC&C	
	Q7/20	Evaluation and assessment of smart sustainable cities and communities	

Underlining indicates revised parts.

IoT: Internet of Things
SC&C: smart cities and communities

with existing open source communities, to examine the status of use of open source programs within ITU-T, and to study the relation between the licensing mechanism and the copyright guideline.

5. Future plan

The second TSAG meeting will be held in Geneva from February 26 through March 2, 2018.

Hideyuki Iwata
Senior Research Engineer, Supervisor, Research and Development Planning Department, NTT.
He received a Ph.D. in electrical engineering from Yamagata University in 2011. From 1993 to 2000, he conducted research on high-density and aerial optical fiber cables at NTT Access Network Service Systems Laboratories. Since 2000, he has been responsible for standardization strategy planning for NTT research and development. He has been a delegate of International Electrotechnical Commission (IEC) Subcommittee 86A (optical fiber and cable) since 1998 and of the ITU-T Telecommunication Standardization Advisory Group since 2003. He is a vice-chair of the Expert Group on Bridging the Standardization Gap in the Asia-Pacific Telecommunity Standardization Program Forum. In 2004, he received an award from the IEC Activities Promotion Committee of Japan for his contributions to standardization work in IEC.

Event Report: NTT Communication Science Laboratories Open House 2017

Tsutomu Hirao, Yasuhiro Takahashi, Yasue Kishino, Masafumi Matsuda, and Shinya Takamuku

Abstract

NTT Communication Science Laboratories Open House 2017 was held in Keihanna Science City, Kyoto, on June 1 and 2, 2017. Nearly 1800 visitors enjoyed 6 talks and 29 exhibits, which focused on our latest research activities and efforts in the fields of information and human sciences.

Keywords: information science, human science, artificial intelligence

1. Overview

At NTT Communication Science Laboratories (NTT CS Labs), we are engaged in two types of fundamental research. The first is research to benefit the future. We are leading the world in preparing for a new era by developing technologies that will bear fruit five to ten years from now and by accumulating knowledge. The second is research that addresses the needs of modern society. We are utilizing the technologies and knowledge we have cultivated to tackle societal challenges directly confronting us now.

NTT CS Labs Open House has been held annually with the aim of introducing the results of the labs' basic research and innovative leading-edge research to both NTT Group employees and visitors from companies, universities, and research institutions who are engaged in research, development, business, and education.

This year, Open House was held at the NTT Keihanna Building in Kyoto on June 1 and 2, and nearly 1800 visitors attended it over the two days. We prepared many hands-on exhibits to allow visitors to intuitively understand our latest research results and to share a vision of the future where new products based on the research and development (R&D) results are widely used. We also organized an invited talk. This article summarizes the event's research talks and exhibits.

2. Keynote speech

Open House started with a speech given by Dr. Eisaku Maeda, Vice President and head of NTT CS Labs (currently, Professor at Tokyo Denki University), entitled "Basic Research in the Age of AI×IoT×BigData—New Design of R&D Processes by the Blending of Science and Engineering," (**Photo 1**).

Photo 1. Dr. Eisaku Maeda delivering keynote speech.

Dr. Maeda pointed out that the convergence of three fundamental technologies—new data technology known collectively as artificial intelligence (AI), new data infrastructure called IoT (Internet of Things), and vast collections of data of unprecedented size and quality referred to as big data—is now having a major transformative impact on the world. He also pointed out that pure science and engineering have always been pursued as independent disciplines, but this is now undergoing a major change caused by developments in the above technologies. Through this change, feedback loops from science to engineering to implementation and back to science occur repeatedly, becoming ever thicker, and the sequence revolves at an ever faster rate. He stated that in these circumstances, even areas of scientific research that up to now have been conducted in idyllic independence must be strategically reexamined in terms of overall economic efficiency.

3. Research talks

Four talks were given, as summarized below, which highlighted recent significant research results and high-profile research themes. Each presentation introduced some of the latest research results and provided some background and an overview of the research. All of the talks were very well received.

(1) "Discovering hidden structure in big data—Knowledge discovery based on probabilistic latent variable models," by Dr. Tomoharu Iwata, Ueda Research Laboratory

Dr. Iwata explained a generative model approach that can automatically find intrinsic latent features from a huge amount of data, which can be easily obtained from the Internet and sensors. He provided guidelines for modeling data by introducing specific models for some applications, such as topic extraction and object matching (**Photo 2**).

(2) "Generative personal assistance with audio and visual examples—Deep learning opens the way to innovative media generation," by Takuhiro Kaneko, Media Information Laboratory

Mr. Kaneko introduced generative personal assistance with audio and visual examples, which can give feedback or instructions to a person who wishes to perform a task better or to do new things. He explained a new deep learning approach to overcome the limitations of existing methods based on manually defined rules and discussed future directions in this research field (**Photo 3**).

(3) "What is special about excellent batters?—The essence of fine skills in the light of implicit brain functions," by Dr. Makio Kashino, Sports Brain Science Project

Dr. Kashino focused on batting in baseball and explained the mechanisms of implicit brain functions, that is, information processing in the brain that the players themselves are not aware of. He presented the results of analyzing various biological data obtained from players and pointed out the key role of implicit brain functions (**Photo 4**).

(4) "Does breathing change the impression of a sound?—Interactive relationship between respiration and sound, and its effect on emotion," by Dr. Takashi G. Sato, Moriya Research Laboratory

Dr. Sato explained the experimental results indicating that breathing (inhalation, exhalation) makes a difference in how the body responds to presented

Photo 2. Dr. Tomoharu Iwata giving research talk.

Photo 3. Takuhiro Kaneko giving research talk.

Photo 4. Dr. Makio Kashino giving research talk.

Photo 6. The latest research results were exhibited.

Photo 5. Dr. Takashi G. Sato giving research talk.

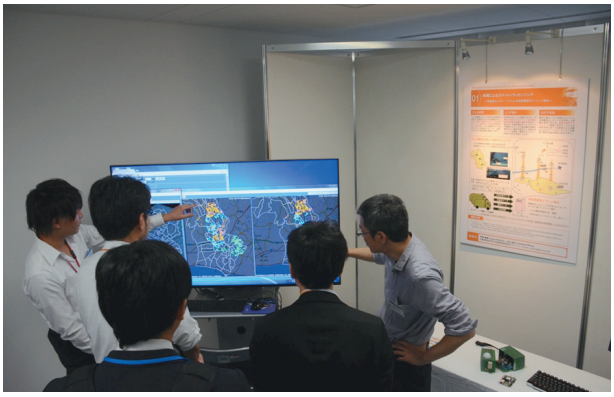


Photo 7. Researchers explaining the research results to visitors.

sound. He also discussed the vision of human interfaces that modify their stimulus presentation according to the state of human activity (**Photo 5**).

4. Research exhibits

Open House featured 29 exhibits displaying NTT CS Labs' latest research results. We categorized them into four areas, "Science of Machine Learning," "Science of Computation and Language," "Science of Media Information," and "Science of Human and Communication." Each exhibit was housed in a booth and employed techniques such as slides on a large-screen monitor or hands-on demonstrations, with researchers explaining the latest results directly to visitors (**Photos 6** and **7**). The following list gives the titles of the research exhibits in each category. More details can be found on the website [1, 2].

4.1 Science of Machine Learning

- Smart city sensing using municipal vehicles—Spatio-temporal city event detection via car-mounted sensors
- Understanding human activity patterns in cities—Spatio-temporal analysis of city activities
- Optimization of real-time collective navigation—Finding efficient navigation by Bayesian optimization
- Inferring pedestrian flow while protecting privacy—Probabilistic behavior model for discovering pedestrian flow
- Stable deep learning for time-series data—Preventing gradient explosion in gated recurrent units
- Fast mining of relationships of large-scale data—An efficient algorithm for L_1-graph construction

4.2 Science of Computation and Language

- Generating absolutely secure shared secret keys—Quantum key distribution using practical devices
- Genuine physical random bit generator—Stably preserving unpredictability originated in entropy sources
- Enumeration of tiling patterns—An efficient algorithm for finding exact covers
- Picture books: a child's first textbook—Relation between picture book corpus and child vocabulary
- Supporting non-native English speakers' listening—Exploring display methods for automated captions
- How to get your favorite translation—Controlling neural machine translation by prefix constraints
- Learning 'neat' semantic representation of words—Neural word embeddings with high usability

4.3 Science of Media Information

- Illumination-based color enhancement—Color enhancement by optimizing illumination spectrum
- Hidden Stereo—Hiding phase-based disparity for viewers without 3D glasses
- Generative personal assistance—Deep learning opens the way to innovative media generation
- Converting English speech to native-like pronunciation—Speech conversion using vocal tract model and deep generative models
- On-time sound delivery by high-quality telephones—CLEAR: low-delay near-lossless codec for audio transmission
- Listening languages around the world simultaneously—Multilingual speech identification & recognition technology
- Personalizing your speech recognizer—Neural network adaptation for automatic speech recognition
- Turn-taking matters in conversation recognition—Robust speech processing using speakers' activity estimation

4.4 Science of Human and Communication

- Does this sound like a native?—Neural mechanisms of speech rhythm perception
- Natural sounds and our auditory system—Reconsidering our auditory system under natural environments
- Reading minds from unconscious eye movements—Decoding implicit mind from fixational eye movements
- Linkage between touch behavior and touch feeling—Estimating tactile perception by analyzing hand motions
- Easier to connect calls during disaster congestion—Control to induce voluntary reduction of call duration
- Brain science and ICT elucidate the athletic brain—Sports brain science based on body-mind reading
- Guide you anywhere by Buru-Navi—Tactile navigation in indoor/outdoor/virtual spaces
- Discussion with AI: City life vs. countryside life—Discussion dialogue system based on argumentation structures

5. Invited talk

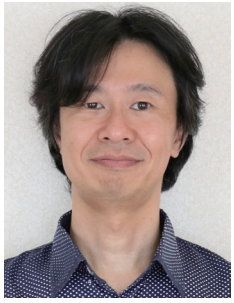
This year's event also featured an invited talk by Mr. Takamasa Sakai, R&D Director, Hakuhodo Institute of Life and Living (a think tank of the Japanese advertising firm Hakuhodo). The title of his talk was "Four scenarios of future city." He described current social issues in Japan such as the aging of the population, unstable employment, and the increase in one-person households, obtained from the results of a public opinion survey on the future of cities. He also talked about living environments and human relationships in the future, taking into account the above issues and the survey results. He then derived four scenarios of future cities—*a city without keys*, *a city without addresses*, *a city without walls*, *and a city without windows*—and explained them from various perspectives, including a technology perspective, by examining the results of a public opinion survey on the four scenarios.

6. Concluding remarks

Just like last year, many visitors came to NTT CS Labs' Open House 2017 and engaged in lively discussions on the research talks and exhibits and provided many valuable opinions on the presented results. In closing, we would like to offer our sincere thanks to all of the visitors and participants who attended this event.

References

[1] Website of NTT Communication Science Laboratories Open House 2017 (in Japanese).
http://www.kecl.ntt.co.jp/openhouse/2017/index.html
[2] Website of NTT Communication Science Laboratories Open House (in English).
http://www.kecl.ntt.co.jp/openhouse/2017/index_en.html

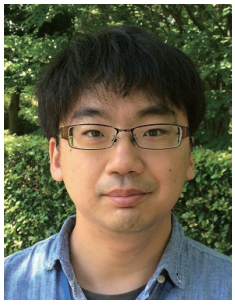


Tsutomu Hirao
Senior Research Scientist, Linguistic Intelligence Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in engineering from the Graduate School of Information Science, Nara Institute of Science and Technology, in 2002. His research interests include natural language processing, machine learning, and artificial intelligence areas.

Yasuhiro Takahashi
Senior Research Scientist, Computing Theory Research Group, Media Information Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in engineering from the University of Electro-Communications, Tokyo, in 2008. His research interests include quantum computing, computational complexity theory, and cryptography.

Yasue Kishino
Senior Research Scientist, Learning and Intelligent System Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
She received a Ph.D. in information science and technology from Osaka University in 2007. Her research interests include ubiquitous computing and sensor networks.

Masafumi Matsuda
Research Scientist, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in human sciences from Hokkaido University in 2004. His resarch area is evolutionary minded social psychology.



Shinya Takamuku
Research Scientist, Sensory and Motor Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in engineering from Osaka University in 2008. His research focuses on the computational aspects of human motor control and perception.

Arkadin Brings Businesses into the Future with New Cloud Unified Communications Services and Digital Operations Strategies

Thomas Valantin, Jean-Pierre Dacher, Melanie Reversat, Nicolas Coudret, Zach Katsof, Hugues Treguier, and Nikki de Kretser

Abstract

As an NTT Communications company, Arkadin is at the forefront of digital cloud communications technologies. We are transforming and adapting rapidly to take advantage of new business opportunities in this fast growing and dynamic market sector. Several services were introduced and a new partnership struck in the past few months, which we believe will help enhance our market leadership and drive greater success for our customers.

Keywords: cloud, unified communications, digital workplace

1. Introduction

As an NTT Communications company, Arkadin is at the forefront of digital cloud communications technologies [1]. This is a fast growing and dynamic market sector that is expected to reach up to €30 billion by 2020—five times more than Arkadin's traditional conferencing market, and we are transforming and adapting rapidly to take advantage of new business opportunities in this area.

Our product and operations teams are constantly searching for innovative technologies and business strategies to ensure our customers are prepared for the demands of a changing digital landscape and a new generation of workers. This is essential if we are to meet our goal of bringing them into the future through simple-to-use cloud collaboration services that enable a true digital workplace for greater productivity and return on investment (ROI).

In the past few months, we have launched services and established noteworthy partnerships that we believe will help enhance our market leadership and drive greater success for our customers.

2. ArkadinVision smooths the transition from conferencing to unified communications

It is commonly known that today's digital workplace requires collaboration tools that are more converged, integrated, and user-centric than those of previous generations. The continued popularity of BYOD (bring your own device), momentum for Skype for Business in enterprises, and the increasing use of video for vertical applications are key factors. As a result, we saw a need for a unified meeting service that would complement and integrate perfectly with our trusted partner solutions from Microsoft, Cisco, Vidyo, and BlueJeans.

ArkadinVision is our answer. Through a partnership with Pexip, we have designed a digital meeting platform to deliver an excellent customer experience that is unique to the market. The user interface offers

the highest level of simplicity and interoperability available today. Customers can easily connect anytime, anywhere, and across diverse endpoints. Whether they join from room systems, Skype for Business, or from a mobile device in one easy click, they will enjoy high quality video, audio, and screen sharing—all with a fresh modern look and feel. Moreover, they will have the convenience of one trusted supplier, a simple license model, and competitive, highly flexible pricing plans.

While our sweet spot is the mid-market, ArkadinVision can support any size business for all their audio, web, and video collaboration needs, especially as they transition from conferencing to unified communications (UC). ArkadinVision also serves as an evolutionary bridge to the digital workplace for our existing ArkadinAnytime and ArkadinAnywhere legacy customers who will be able to streamline multiple collaboration solutions under one license for added value and cost savings. For businesses that already have UC, ArkadinVision can supplement important features that may be inadequate or missing from their solutions.

Our guiding principle in development was to avoid common customer pain points that can lead to low levels of adoption. To circumvent this problem, ArkadinVision offers:

- A simple interface to make meetings more productive and enjoyable
- A limited user footprint that supports plug-in free Web Real-Time Communications (WebRTC)* browsers
- High quality high-definition (HD) audio on voice over Internet protocol (VoIP) and PSTN (public switched telephone network) dial in/out
- Pexip's highly available and scalable platform for high levels of performance and downtime
- Comprehensive training and onboarding programs

Customers participating in our trials are enjoying significant benefits that are driving high value collaborative experiences:

- Of customers trying our system, 47% have achieved high levels of adoption, compared to the 20–25% adoption range that is typical for most video meeting services.
- An impressive 97% of respondents have found the service easy to use, with many commenting on the high quality of the video.

ArkadinVision is available in Europe, North America, and Latin America and will be launched in Asia later this year.

Link to press release: https://www.arkadin.co.uk/arkadin-unveils-arkadinvision-new-digital-meeting-space-simplicity-affordability
Link to video: http://vidyard.arkadin.com/watch/v42XspEt5aK1pFMX9hxTHM

3. New Cisco services are tapped to give customers greater choice in UC and collaboration

As a technology agnostic company, we compete by offering customers a wide array of products for solving their unique cloud UC needs and supporting them with our trademark global, yet local approach to service. Cisco Spark is the latest addition to our full complement of UC services. This new era collaboration tool provides messaging, meeting, and calling services directly from the Cisco Collaboration Cloud. The highly secure, scalable, and easy to use service enables users to message, meet, or call anyone, anywhere, at any time, and from any device. Whether participants join from the new generation Cisco Spark Board, a room system, a desktop phone, or a mobile device, they all have the same meeting experience. Through a single click, they can turn a phone call into a video meeting, and with a simple swipe, a video call can be moved from a room system to a mobile phone and to another room system. Cisco Spark is available in a monthly, per user subscription service.

Cisco Spark is a new team collaboration application that lets teams work in virtual rooms. Within these rooms, team members can send instant messages (either in groups or one-to-one), share files, and make video calls. The solution called Cisco Cloud Connected Audio – Service Provider (CCA-SP), which has been branded Cloud Connected Audio – Arkadin, has also been added to our offering for optimizing WebEx meetings as a fully integrated solution offering audio, web, and video. It offers significant advantages: a predictable cost model, a hybrid audio environment, HD VoIP; the ability to leverage existing customer IP telephony networks, and a secure, scalable global cloud infrastructure. Arkadin is among the limited number of Cisco partners to offer and implement CCA-SP with a WebEx web conferencing platform.

All Cisco Spark and Cloud Connected Audio customers benefit from our locally administered, highly

* WebRTC: A standard of communications protocols and application programming interfaces that enable web browsers to use real-time voice, text, and video communications.

personalized support strategy, which gives Arkadin a compelling advantage in the industry.

4. Arkadin Managed Services for Skype Cloud Voice guides organizations through the voice transformation journey to achieve greater ROI with Microsoft Office 365

As Microsoft continues to roll out advanced Skype cloud voice capabilities, organizations are looking for a trusted voice partner with the expertise and experience to navigate the voice transformation journey. In close partnership with Microsoft, we are providing our newly launched Arkadin Managed Services for Skype Cloud Voice to help customers translate their cloud voice strategy into an action plan for success, while mitigating risk and the all too common potholes in voice transformation projects. The Microsoft Skype Operations Framework certified solution addresses planning, readiness, deployment, user enablement, and ongoing managed services. Customers benefit from cost and risk controls, simplified administration, increased security, an improved user experience and a highly skilled transformation team to guide users through the voice transformation journey. Arkadin announced the new service in conjunction with the Microsoft Ignite show in Orlando, Florida, in September 2017.
Link to press release: https://www.arkadin.com/about-us/news-media/arkadin-announces-new-managed-services-office-365-enable-businesses-successfully
Link to video: https://youtu.be/zihSFJKJtmA

5. Arkadin and Unify Square partner will accelerate global Skype for Business adoption

A new partnership with Unify Square, the leading provider of 24x7 cloud managed services for Skype for Business, will enable Arkadin to be the first UCaaS (UC-as-a-Service) provider to put user quality of experience at the heart of its platform. Arkadin Total Connect Enterprise, a set of business applications and services around Skype for Business across Microsoft Office 365 and Arkadin's global private cloud, will embed Unify Square's PowerSuite software and cloud managed services for ensuring exceptional monitoring and optimization.

"Quality of experience for the end user remains the cornerstone of a successful unified communications deployment in the enterprise world. Embedding PowerSuite into our platforms further elevates our focus on delivering, and maintaining, a superior user experience," said Christophe Reyes, Managing Director, Unified Communications Service Line for Arkadin. He added, "Unify Square's exceptional monitoring and optimization of Skype for Business will ensure our users stay at the heart of our partnership."
Link to press release: https://www.arkadin.co.uk/about-us/news-media/arkadin-and-unify-square-partner-accelerate-global-skype-business-adoption

6. SmartStart is a game changer in solving user adoption headaches for UC services

One of the long-standing challenges of selling online collaboration tools to enterprises is user adoption. Traditional onboarding takes as long as four weeks. Arkadin set out to significantly shorten this time to minimize the disruption caused by business changes and to get users up and running quickly and efficiently. The problem was tackled through the creation of a best-in-class end-user product onboarding portal called SmartStart. The objective was to give end users a more personalized experience and far more control over the process, which is not possible with many traditional onboarding strategies.

The easy and intuitive four-step portal has everything required for the onboarding journey in one easily accessible place that can be accessed from any mobile device and customized based upon each user's personal preferences. It is powered by Oracle Eloqua, Vidyard, and the open source Orchard CMS (content management system) (**Fig. 1**).

Vidyard enables us to see which videos stored across the SmartStart microsite are being consumed by our customers, which sections of the videos they are finding the most valuable, and the identities of the people who have been watching the videos (**Figs. 2** and **3**). Because video assets are the most expensive asset used with the SmartStart program, the use of Vidyard is key for us to understand which videos are most popular and which format the users engage with the most so that we can ensure that future videos are created in the most cost-effective manner for maximum ROI.
Link to SmartStart for ArkadinVision: https://smartstart.arkadin.com/vision/home

7. Metrics show onboarding is reduced from 4 weeks to less than 10 minutes

Beta customers who navigated through to the SmartStart site were measured to determine their

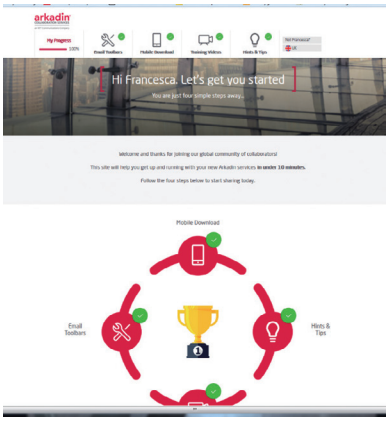


Fig. 1. A *splash screen* greets the user with a personalized greeting and a progress display. It includes a progress indicator in the top navigation bar as well as the "trophy" status.

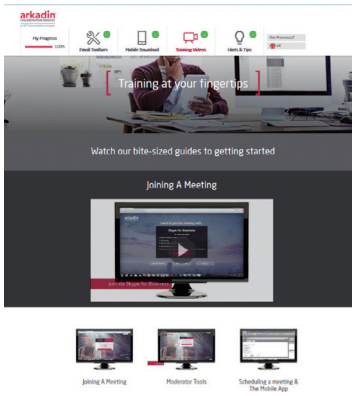
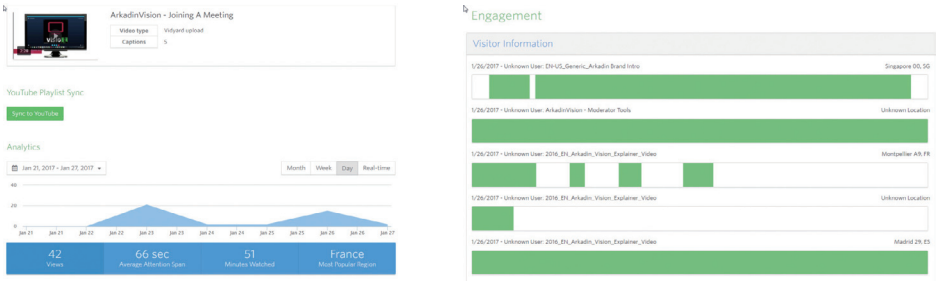


Fig. 2. Training videos are hosted in Vidyard and embedded into the SmartStart portal. Viewing information is tracked and sent to Eloqua. When all videos have been watched to completion, the status is updated in Eloqua and reflected across the portal.



(a) (b)

Fig. 3. (a) Screen shot of video engagement over time and (b) user-specific video consumption. User behavior information is shown here with an anonymous user's details for privacy reasons.

completion rates. The site average was 5.21 pages per session, which is very good. The metrics show a high level of engagement with the site, with users leaving the site on average after 5 minutes 45 seconds. This means that the users are engaging with the site and taking what they need from their visit, and it proves that we can reduce the onboarding period of the end users from 4 weeks to less than 6 minutes on average.

In the beta trials for ArkadinVision, adoption rates toppled industry norms. The average end-user adoption rate reached 40%, which is significant considering the average for similar video conferencing services is a paltry 24%, and Arkadin users were able to onboard in less than 10 minutes. SmartStart played a critical role in achieving these results.

It is not just the end users who are benefiting. Customers tell us their information technology (IT) departments are not as burdened with support questions since their employees can be directed to the portal for questions. Thus, IT decision makers are experiencing a more immediate ROI from their purchase. In addition, it will ultimately make it possible for Arkadin to reduce the amount of onboarding communications that we send to our customers via email by enabling the microsite to replace these emails. This is a major step forward for Arkadin, as it enables us to manage over-communication by email with an efficient tool that still has Eloqua as its backbone.

The portal is offered in Arkadin's new ArkadinVision digital meeting service [2] and in the company's other proprietary services, including ArkadinAnytime (audio) [3] and ArkadinAnywhere (web conferencing) [4].

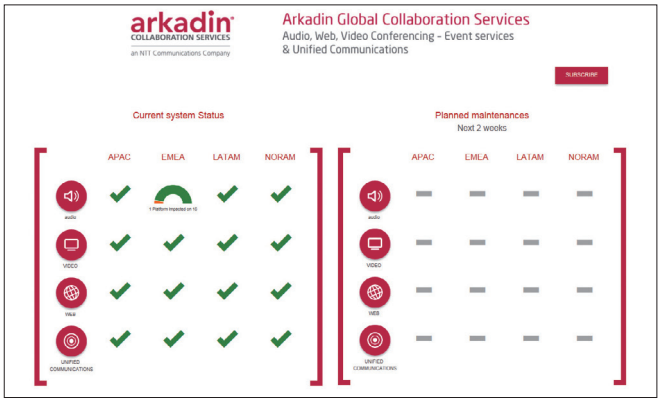


Fig. 4. Screen shot of online customer notification portal.

Link to press release: https://www.arkadin.com/about-us/news-media/arkadin%E2%80%99s-smartstart-game-changer-solving-user-adoption-headaches-unified

8. Achieving world-class operations drives our customers' digital transformation journey

To ensure our customers experience the best cloud communications services and a seamless user experience, we are focusing on three operations initiatives:

(1) Alignment with the expectations of our customers

Findings from our semi-annual customer survey are used in program development to ensure client needs are being addressed. Current priorities include a focus on voice quality and customer communications. We monitor the quality of audio streams to gain a continuous view of customers' voice experiences on a global basis, which helps us to optimize voice traffic in conjunction with carriers. Our online customer notification portal enables us to proactively notify customers on the status of our infrastructure and any service issues in real time. For example, the report details any recent outages (**Fig. 4**).

(2) Key performance indicators and measurement

Our key performance indicators and measurement tools identify how well we delivered on the requested levels of performance. Arkadin launched a 'four 9s' program to establish a bar for availability and downtime targets on infrastructure for our internally developed services such as ArkadinAnytime (audio), ArkadinAnywhere (web conferencing), and ArkadinVision (digital meetings).

(3) Product development

We have adopted development and operations (DevOps) principles to enable a more integrated and agile way of bringing our products to market. This powerful methodology for product development will ensure we have the most efficient workplace culture and processes in place to build, test, and release our services rapidly and reliably, and that we are well positioned to respond to customer feedback. ArkadinVision was launched using DevOps for the first wave of this program. In the second wave, we will integrate more DevOps activities and further automate testing and deployment processes.

9. Digital cloud communications services and digital operations strategy results in prestigious awards

Our strategy for bringing our global clients into the future with simple-to-use cloud collaboration services that enable a true digital workplace for greater productivity and ROI is resulting in prominent awards.

(1) Frost & Sullivan Asia Pacific ICT Awards

Arkadin was recognized as the 2017 Frost & Sullivan Asia Pacific Collaboration Service Provider of the Year for the fifth consecutive time (**Fig. 5**). In commenting on the award, Zi Ning Chong, Research Analyst, ICT Practice - Asia Pacific, at Frost & Sullivan, said: "In a highly competitive and commercialized market, Arkadin maintained its customer-oriented strategy and continued its profitable revenue growth in 2016. To strengthen the interoperability between collaboration technologies, the company further elevated the strategic partnerships with Microsoft and Cisco, notably with the launch of

Arkadin Total Connect, providing a fully integrated UC ecosystem with voice-enabled Office 365, Cisco CCA-SP, and Cisco Spark solutions. To stand out in the crowd of collaboration service providers, Arkadin's diversified UC service portfolio and direct-touch Go-to-Market strategy helped the company to further improve its customer retention rate and win new customers across the region. Also, we expect the latest digital meeting space, ArkadinVision, which is being launched in the Asia Pacific region this year, will offer an impetus for the growth of streamlined collaboration in the near future."

Link to press release: https://www.arkadinapac.com/about-us/news-media/ntt-communications-and-arkadin-honored-frost-sullivan-asia-pacific-ict-awards

Fig. 5. 2017 Frost & Sullivan Asia Pacific Collaboration Service Provider of the Year.

(2) Communications Solutions Product of the Year 2017

ArkadinVision was recognized for exceptional innovation by global integrated media company TMC (**Fig. 6**). "Congratulations to Arkadin for being honored with a Communications Solutions Product of the Year Award," said Rich Tehrani, Chief Executive Officer, TMC. "ArkadinVision is truly an innovative product and is amongst the best solutions brought to market in the past 12 months that facilitate business transformation in voice, data, and video communications."

Link to press release: https://www.arkadin.co.uk/about-us/our-awards/arkadin-vision-digital-meeting-space-receives-prestigious-industry-awards

Fig. 6. TMC Communications Solutions Product of the Year 2017.

(3) 'Best-in-Biz International' Award

Arkadin's newly launched SmartStart end-user onboarding portal received a Gold award in the Best New Product Feature of the Year category from this prominent global awards organization (**Fig. 7**). "We created SmartStart in partnership with Oracle Eloqua to create an onboarding journey that is fully automated and personalized to help improve our customers' experience in solving the vexing problem of low user adoption," says Nikki de Kretser, Vice President, Marketing, at Arkadin. "SmartStart not only helps users to get started quickly and easily, it removes the hassle faced by the IT group tasked with embedding these new digital transformation technologies. We're thrilled to have this recognition for our efforts."

Link to press release: https://www.arkadin.co.uk/about-us/our-awards/arkadin-vision-digital-meeting-space-receives-prestigious-industry-awards

Fig. 7. Best-in-Biz International Award.

(4) Stratus Awards for Cloud Computing

Arkadin is recognized as a visionary and innovator leading the cloud revolution in collaboration for enterprises, with offerings that are differentiated and uniquely adding value to our daily lives (**Fig. 8**).

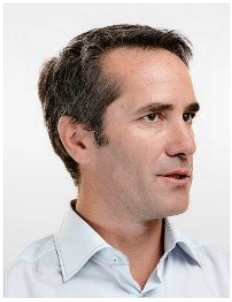
Link to press release: https://www.arkadin.co.uk/about-us/our-awards/arkadin-named-global-leader-2017-stratus-awards-cloud-computing

Fig. 8. Stratus Award for Cloud Computing.

References

[1] D. Jaubert, "Arkadin: Global Unified Communications & Collaboration Champion in a Dynamic Digital Workplace," NTT Technical Review, Vol. 14, No. 9, 2016. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201609ga1.html
[2] Website of ArkadinVision, https://www.arkadin.co.uk/services/video-conferencing/arkadinvision
[3] Website of ArkadinAnytime, https://www.arkadin.com/solutions/audio-conferencing
[4] Website of ArkadinAnywhere, https://www.arkadin.com/solutions/web-conferencing

Trademark notes



Thomas Valantin

Chief Commercial and Transformation Officer, Arkadin.

In his role as Chief Commercial & Transformation Officer (CCTO), Thomas Valantin is responsible for Arkadin's transformation program and global sales strategy and sales operations. He joined Arkadin in November 2011 following a career in international sales management with high-tech and telecom companies in Europe and Latin America, which included 12 years in the video conferencing industry at Polycom, Tandberg, and Cisco Systems. He graduated from the International Business School of Paris Chamber of Commerce and holds an MBA from ESIC Madrid in International Business Trading.

Jean-Pierre Dacher

Chief Technology and Information Officer, Arkadin.

Jean-Pierre joined Arkadin as Chief Technology & Information Officer (CTIO) in 2016 and is responsible for leading product, development, operations, and information systems. He has 30 years of international business experience that includes roles in engineering planning, and product and project management with enterprises such as Hewlett Packard, SAP, Business Objects, and Murex, where he was chief operating officer. He holds an engineering degree from Ecole Nationale d'Electronique et Radio-Electricité and a Mastere from Ecole Nationale Supérieures des Télécommunications.

Melanie Reversat

Senior Product Manager for Digital Meeting Spaces, Arkadin.

As Senior Product Manager for Digital Meeting Spaces, Melanie is responsible for overseeing the ArkadinVision, Adobe Connect, BlueJeans, and Vidyo product lines. Melanie joined Arkadin in April 2013. She holds a master's degree in Statistics from Paris XI University.

Nicolas Coudret

Head of Cisco Unified Communications Strategy, Arkadin.

In his role as Head of Product Management Cisco, Nicolas oversees product management for Cisco's full suite of Cloud Collaboration solutions at Arkadin, including WebEx, CCA-SP, and Spark. Nicolas joined Arkadin in July 2015. He holds a master's degree in video broadcasting, multimedia, and distance learning systems from Paris University.

Zach Katsof

Director of Unified Communications, Americas, Arkadin.

As Director of Unified Communications, Americas, Zach is leading Arkadin's UC efforts for the Americas. He joined Arkadin in June 2006. Zach has a bachelor's degree in political studies from Queen's University, Kingston, Ontario and an MBA from the Georgia Institute of Technology.

Hugues Treguier

Director of Strategy & Products – Unified Communications, Arkadin.

In his role as Director of Strategy & Products – Unified Communications, Hugues is responsible for Arkadin's global UC strategy and product management, including building Arkadin's UC vision and strategy, managing the construction of Arkadin's UCaaS service offers portfolio, overseeing the evolution of the UCaaS roadmap and ensuring its timely delivery, and managing the UC partner ecosystem. He graduated from the Ecole d' Ingenieurs des Technologies de l' Information et du Management. He joined Arkadin in October 2008.

Nikki de Kretser

Vice President of Marketing, Arkadin.

In her position as Vice President of Marketing, Nikki is responsible for overseeing all global marketing activities to support revenue growth and Arkadin's business strategy to be a leading cloud communications player in North America, Latin America, Europe, and Asia. Nikki joined Arkadin in October 2014. She is a graduate of the University of Newcastle-upon-Tyne and holds a master's degree in marketing.

World's Largest Transmission Capacity with Standard Diameter Multi-core Optical Fiber— Accelerated Multi-core Fiber Application Using Current Standard Technology

1. Introduction

NTT and six partners, KDDI Research, Inc. (KDDI Research), Sumitomo Electric Industries, Ltd. (Sumitomo Electric), Fujikura Ltd. (Fujikura), Furukawa Electric Co., Ltd. (Furukawa), NEC Corporation (NEC), and Chiba Institute of Technology (CIT) have demonstrated the world's largest transmission capacity of 118.5 Tbit/s using a multi-core fiber with four optical paths (cores) with the same diameter as the optical fiber in current use.

Having a conventional glass diameter (125 μm) in accordance with the international standard enables us to effectively use existing optical fiber fabrication and optical connector technologies. This achievement proves that the concept of a multi-core fiber based long-haul and large capacity transmission system consisting of multiple vendor technologies is viable, and it makes significant progress in the practical use of multi-core fiber technology (**Fig. 1**).

Our objective is to introduce the standard diameter multi-core fiber by the early 2020s. We will also continue to contribute to developing a future optical infrastructure that can support various data communication demands.

This remarkable achievement was reported in early August as a postdeadline paper at the Opto-Electronics and Communications Conference (OECC 2017), the largest conference on optical communication in Asia Pacific-Rim, which was held at the Sands Expo and Convention Centre, Singapore, from July 31 to August 4, 2017. This work was partially based on work commissioned by the National Institute of Information and Communications Technology (NICT).

2. Research background

The worldwide spread of various mobile terminal and data services has led to a continuous increase in transmission capacity of more than 10% per year all over the world. However, the increase in capacity may not keep up with the increase in demand, a trend that may cause a capacity crunch in the currently used optical fiber by the late 2020s. Moreover, the continued expansion of the optical fiber count (i.e., the number of fibers used) and the convergence of optical wiring—particularly in datacenters and/or central offices—which is caused by the worldwide data capacity increase, would be a serious problem.

Against this background, multi-core fiber having multiple optical paths (cores) in one fiber has been investigated intensively all over the world in order to overcome the future capacity crunch and to achieve high density or space-saving optical facilities. For example, ultralarge capacity transmission experiments using a multi-core fiber with ten cores or more have been demonstrated. However, this high core count multi-core fiber usually needs a thicker glass diameter, and it requires an extreme advance in the fabrication process and further development of subcomponents. Consequently, about ten years or so is

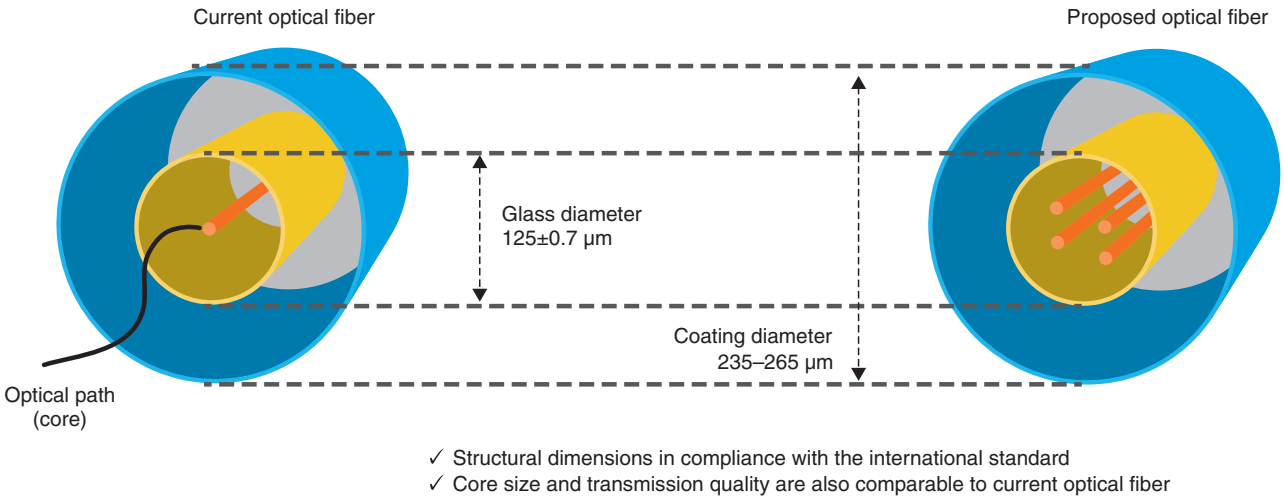


Fig. 1. Features of proposed multi-core fiber.

expected to be necessary to make the high core count multi-core fiber practical.

To accelerate the use of multi-core fiber technology, NTT, KDDI Research, Sumitomo Electric, Fujikura, Furukawa, NEC, and CIT developed a multi-core fiber with a conventional diameter in accordance with the current international standard. This makes it possible to use existing optical fiber technology even though the number of cores is limited to four or five.

3. Design guidelines

The fabrication process of an optical fiber starts with preparing a relatively large glass rod with a diameter of several to ten centimeters called a preform. An optical fiber is made by melting and drawing the preform while maintaining the geometrical similarity. When the glass diameter of an optical fiber is doubled (e.g., from the conventional 125 μm to 250 μm), the fabrication length obtained with the same size preform is reduced to a quarter. Therefore, the increase in the glass diameter directly affects the mass productivity of an optical fiber. The current optical communication system commonly uses a single-mode fiber (SMF) with a core diameter of about 10 μm, which can be used in the entire telecommunication wavelength region (1260–1625 nm).

We therefore aim to achieve a multi-core fiber with two features: i) respective glass and coating diameters of 125 ± 0.7 μm and 235–265 μm in compliance with the international standard of the current optical fiber, and ii) a fiber whose individual core has a similar transmission quality to that of the commonly used SMF. In a multi-core fiber, the optical signal interference between neighboring cores should be reduced sufficiently. NTT and KDDI Research revealed that four to five cores can be arranged in a 125-μm glass diameter.

Our vision of future communication traffic sees the widespread distribution of high-resolution images and video such as 4K/8K and the full-scale popularization of machine-to-machine services. The obvious assumption is that capacity demands will skyrocket. Thus, NTT has been actively pursuing a wide spectrum of studies into achieving ultralarge capacity transmission of signals exceeding 100 Gbit/s. Furthermore, with the recent increase in traffic between datacenters, greater capacity must be matched by greater cost savings.

Greater capacity is most directly achieved by using multiple bands (C band + L band etc.), which expands the usable bandwidth of the optical fiber. However, several problems arise if we attempt to transmit multiband signals by wavelength division multiplexing (WDM) over dispersion shifted optical fiber (DSF: dispersion shifted fiber) cable. The zero dispersion wavelength of DSF lies in the extended band (C band), and around this wavelength, the signal is degraded by nonlinear effects (especially four-wave mixing). The common solution, unequally spacing the signal wavelengths in the C band, degrades the wavelength utilization efficiency. This is a barrier to further capacity increases.

4. Multi-core transmission line composed of multiple vendor optical fibers

Sumitomo Electric, Fujikura, and Furukawa used the above design guidelines to individually fabricate multi-core fibers with four cores over 100 km in length. All the multi-core fibers developed by these companies can be used in the 1260–1625 nm wavelength region, and they have similar transmission properties to the current SMF. (For comparison, mode field diameter (MFD) at 1550 nm is 9–10 μm.)

The fabricated multi-core fibers were divided into segments 20–40 km long, and three transmission spans with a length of 104–107 km were reconstructed by splicing the multi-core fibers provided by different vendors. Satisfactory low loss characteristics comparable to the conventional SMF were achieved. The average loss of four cores in each span was 0.22 dB/km or less, including splicing losses. We applied fusion splicing to splice two multi-core fibers by melting each end. A 0.21-dB/km loss property was achieved as the average of all three spans.

These achievements indicate that our standard diameter multi-core fiber with similar transmission quality (MFD) to that of the conventional SMF will make it possible to greatly improve the productivity of the multi-core fiber made by effectively using the existing fabrication technology and knowledge.

5. Beyond 100-Tbit transmission

A multi-core transmission system was constructed by concatenating three spans. Three multi-core optical amplifiers fabricated by NEC, KDDI Research, NTT, and Furukawa were inserted at each end of three spans in order to compensate for the signal attenuation. Cladding pumping type multi-core optical amplifiers that were expected to reduce power consumption were used, and a 16% improvement was confirmed in this achievement. To confirm the capability of the constructed multi-core transmission line to beyond 100-Tbit/s transmission, 16QAM (quadrature amplitude modulation) based 116-wavelength division multiplexing signals were prepared, and the output signal quality after 316-km-long transmission was examined.

Fan-in/fan-out devices fabricated by NTT and Furukawa were used to input/output signals to/from each core of the multi-core fiber. Pluggable optical connectors with existing MU (miniature universal coupling)-type or SC (single fiber coupling)-type interfaces, fabricated by CIT and NTT, were used to connect the input/output end of the multi-core transmission lines and fan-in/fan-out devices. These optical connectors have rotational alignment features in order to connect the facing four cores correctly. Thus, the low loss and pluggable optical connection of multi-core structure was achieved.

Satisfactory transmission quality was confirmed in all cores and at all wavelengths. This result is the world's largest transmission capacity of 118.5 Tbit/s for a standard diameter optical fiber. These achievements reveal that multi-core fiber with a standard diameter can be used to achieve an ultralarge capacity transmission system to overcome the capacity crunch in the current SMF.

For Inquiries
Public Relations, NTT Information Network Laboratory Group
Email: inlg-pr@lab.ntt.co.jp
http://www.ntt.co.jp/news2017/1708e/170808b.html

External Awards

2016 LOIS Outstanding Service Award
Winner: Manabu Okamoto, NTT Media Intelligence Laboratories
Date: May 11, 2017
Organization: The Technical Committee on Life Intelligence and Office Information Systems (LOIS), Institute of Electronics, Information and Communication Engineers (IEICE) Information and Systems Society

For his significant contribution to the operation and development of LOIS.

Presidential Citation
Winner: Munekazu Date, NTT Media Intelligence Laboratories
Date: May 23, 2017
Organization: The Society for Information Display

For his outstanding service as Executive Chair of the 2016 International Display Workshops (IDW).

Niwa & Takayanagi Achievement Award
Winner: Shohei Matsuo, NTT TechnoCross Corporation; Yukihiro Bandoh and Seishi Takamura, NTT Media Intelligence Laboratories
Date: May 26, 2017
Organization: The Institute of Image Information and Television Engineers

For their pioneering research on a video encoding method and contribution to international standardization.

IPSJ Yamashita SIG Research Award
Winner: Haruno Kataoka, NTT Service Evolution Laboratories
Date: August 24, 2017 (Award ceremony to be held in March 2018)
Organization: Information Processing Society of Japan (IPSJ)

For "Dynamic Guide Signs Control Pedestrians of Public Facilities."
Published as: H. Kataoka, K. Hashiguchi, K. Wago, Y. Ichikawa, and H. Tezuka, "Dynamic Guide Signs Control Pedestrians of Public Facilities," IPSJ SIG Technical Report, Vol. 2016-UBI-50, No. 13, May 2016.

IPSJ Yamashita SIG Research Award
Winner: Tomohiro Kokogawa, NTT Secure Platform Laboratories
Date: August 24, 2017 (Award ceremony to be held in March 2018)
Organization: Information Processing Society of Japan (IPSJ)

For "Visualization and Study of Incident Response Capability of Organizations Based on ISO 22320."
Published as: T. Kokogawa, Y. Maeda, A. Amano, and Y. Kohno, "Visualization and Study of Incident Response Capability of Organizations Based on ISO 22320," IPSJ SIG Technical Report, Vol. 2017-GN-101, No. 23, Mar. 2017.

Best Paper Award
Winner: Akinori Hosoyamada and Kazumaro Aoki, NTT Secure Platform Laboratories
Date: September 1, 2017
Organization: The 12th International Workshop on Security (IWSEC 2017)

For "On Quantum Related-key Attacks on Iterated Even-Mansour Ciphers."
Published as: A. Hosoyamada and K. Aoki, "On Quantum Related-key Attacks on Iterated Even-Mansour Ciphers," Advances in Information and Computer Security—Proc. of IWSEC 2017, Hiroshima, Japan, Aug./Sept. 2017, pp. 3–18, in Security and Cryptology Series, Vol. 10418, Springer, 2017.

Young Scientist Presentation Award
Winner: Kenta Takata, NTT Basic Research Laboratories
Date: September 5, 2017
Organization: The Japan Society of Applied Physics

For "Controllable One-dimensional Photonic Topological Phase with PT-symmetry Breaking."
Published as: K. Takata and M. Notomi, "Controllable One-dimensional Photonic Topological Phase with PT-symmetry Breaking," The 64th Spring Meeting, 15p-E205-5, Yokohama, Kanagawa, Japan, Mar. 2017.

Papers Published in Technical Journals and Conference Proceedings

Analysis of Inversely Proportional Carrier Sense Threshold and Transmission Power Setting

K. Yamamoto, X. Yang, T. Nishio, M. Morikura, and H. Abeysekera

Proc. of the 14th Annual IEEE Consumer Communications & Networking Conference (CCNC 2017), pp. 13–18, Las Vegas, NV, USA, January 2017.

In this paper, an asymptotic analysis of the inversely proportional setting (IPS) of carrier sense threshold (CST) and transmission power in densely deployed wireless local area networks (WLANs) is presented. In densely deployed WLANs, CST adjustment is a crucial technology to enhance spatial channel reuse, but it can starve surrounding transmitters due to an asymmetric carrier sensing relationship. In order for the carrier sensing relationship to be symmetric, the IPS of the CST and transmission power is a promising approach, i.e., each transmitter jointly adjusts the CST and transmission power in order for their product to be equal to those of others. By assuming that the set of potential transmitters follows a Poisson point process, the impact of the IPS on throughput is formulated based on stochastic geometry in two scenarios: an adjustment of a single transmitter and an identical adjustment of all transmitters. The asymptotic expression of the throughput in dense WLANs is derived and an explicit solution of the optimal CST is achieved as a function of the number of neighboring potential transmitters and signal-to-interference power ratio using approximations. This solution was confirmed through numerical results, where the explicit solution achieved throughput with a loss of less than 8% compared to the numerically evaluated optimal solution.

Starvation Mitigation for Densely Deployed WLANs through Distributed Channel Selection: Potential Game Approach

B. Yin, S. Kamiya, K. Yamamoto, T. Nishio, M. Morikura, and H. Abeysekera

Proc. of the 14th Annual IEEE Consumer Communications & Networking Conference (CCNC 2017), pp. 548–553, Las Vegas, NV, USA, January 2017.

A potential game based distributed channel selection scheme is proposed in this paper to mitigate the flow-in-the-middle (FIM) throughput starvation problem that frequently occurs in dense wireless local area networks (WLANs). The FIM throughput starvation occurs when neighbors of a given node are not within the carrier sense ranges of each other. Since they spatially reuse the channel and at least one of them transmits with a high probability, the node in the middle would detect the channel being occupied for a prolonged time and therefore experience extremely low throughput. The basic idea of the proposed scheme is to let each access point (AP) select the channel that reduces the number of three-node chain topologies on its two-hop neighborhood contention graph. The proposed scheme is proved to be a potential game, i.e., the proposed scheme is guaranteed to converge. Graph-based simulation shows that starvation occurs on 20% of nodes when nodes randomly select their frequency channels. The proposed scheme significantly reduces the number of starved nodes along with iterations, outperforming the compared traditional potential game based scheme.

Manipulation of Self-folded Cell-laden Micro-rolls

T. Teshima, H. Nakashima, Y. Ueno, S. Sasaki, C. S. Henderson, and S. Tsukada

Proc. of the International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS) 2017, pp. 5–7, Montréal, Canada, July 2017.

In this paper, we demonstrated a cell-handling thin polymer film, termed a micro-roll, for encapsulating and manipulating adherent cells. The micro-rolls consisted of double layered films with silk fibroin hydrogel and parylene. The geometrically controlled strain of the films achieved self-folding into three-dimensional (3D) tubular architectures with controllable diameter. Furthermore, a release of the sacrificial hydrogel layer with chelating agents provided high biocompatibility; thereby, multiple cells could be wrapped in the individual micro-rolls. We demonstrated that the embedded cells within the micro-rolls were artificially reconstructed into hollow or fiber-shaped tissue-like structures without cytotoxicity. The cell-laden micro-rolls were selectively collected and freely manipulated with a micro-capillary. This system could potentially provide the mobile templates for bio-interfaces such as the reconstruction of functional tissues and implantable tissue grafts.

Similarity Calculation Method for Binary Executables

A. Nakajima

Proc. of Dagstuhl Seminar 17281, Dagstuhl, Saarland, Germany, July 2017.

This talk first gives an overview of the main ideas, challenges, and the major research papers in this area. Then we introduce our research on a method that can identify the similar function in two given binary executables, even the target binary executables that have some modifications. Lastly, we introduce the state-of-the-art research in this area, and discuss how it can be applied in today's malware analysis.

Immersive Telepresence Technology "Kirari!"

H. Takada

Journal of the Imaging Society of Japan, Vol. 56, No. 4, pp. 366–373, August 2017 (in Japanese).

To achieve natural telecommunication, we have been researching an immersive live experience system using several technologies. It reproduces a variety of events with an ultra-high realistic sensation modality using image, high-fidelity audio, and media transport technology. Our research results enabled us to propose an immersive telepresence concept called "Kirari!" for providing immersive live experiences and to develop a life-size "Kirari!" prototype system. Experiments were conducted to evaluate the influence of content expression from reality and experience in a platform for the "Kirari!" system. This result indicated that "Kirari!" is suitable for large-scale public viewing.

Full Parallax Visually Equivalent Light Field 3D Display Using Linear Blending

M. Date, H. Fujii, and H. Kimata

Proc. of the 9th International Conference on 3D systems and Applications (3DSA 2017), Digest version, p. 521, Busan, South Korea, August 2017.

Linear blending is a powerful method to interpolate images when disparities between the images are small enough. We have developed a real-time highly realistic video communication system using a multiple camera array and glasses-type 3D display with head tracking.

TwinCam: Omni-directional Stereoscopic Live Viewing Camera for Reducing Motion Blur during Head Rotation

K. Tashiro, T. Fujie, Y. Ikei, T. Amemiya, K. Hirota, and M. Kitazaki

Proc. of ACM SIGGRAPH 2017 Emerging Technologies (SIGGRAPH'17), p. 24, Los Angeles, CA, USA, July/August 2017.

We developed an omni-directional stereoscopic live viewing camera (TwinCam) system to reduce the motion blur and latency during head rotation of a remote user wearing a head mounted display (HMD). The TwinCam system consists of two omni-directional live cameras (THETA S, Ricoh), rotation mechanisms with a motor, an image control PC, and an HMD. The camera base rotates synchronously with the azimuth angle of the HMD that the observer is wearing, while each camera lens is at a constant azimuth angle. This camera configuration greatly reduces image flow on the CMOS image sensor in the camera, and eventually, the motion blur on the HMD screens when the HMD rotates. The apparent image latency during the head rotation is minimized by the buffered image. A user study demonstrated that both reduced motion blur and compensated latency were effective in reducing the virtual reality (VR) sickness symptoms.

100-year History and Future of Network System Technologies in Japan

H. Tode, K. Kawashima, and T. Ito

IEICE Trans. Commun., Vol. E100-B, No. 9, pp. 1581–1594, September 2017.

Telecommunication networks have evolved from telephony networks to the Internet, and they sustainably support the development of a secured, safe, and comfortable society. The so-called "switching technology" including the evolved "network system technology" is one of the main infrastructure technologies used for realizing information communication services. On the occasion of the completion of 100 years since the establishment of the Institute of Electronics, Information and Communication Engineers (IEICE), we summarize the history of network system technologies and present their future direction for the next generation. We mainly focus on a series of technologies that evolved through the discussions of the IEICE technical committees on switching engineering, launched 50 years ago, switching systems engineering, and network systems in action.

Evolution and Future of Information Networks

T. Asami, K. Yamaoka, and T. Kishida

IEICE Trans. Commun., Vol. E100-B, No. 9, pp. 1595–1605, September 2017.

This paper looks at the history of research in the Technical Committee on Information Networks from the time of its inception to the present and provides an overview of the latest research in this area based on the topics discussed in recent meetings of the committee. It also presents possible future developments in the field of information networks.
