NTT DOCOMO

Technical Journal

vol.19 No.4 | Apr.2018

DOCOMO Today

- Toward the Co-creation of New AI Services

Technology Reports

- Improving Customer Satisfaction and Operator Efficiency in Call Centers Using AI
  - Speech Recognition IVR -
- Deep Learning Platform Capable of Rapid Creation and Deployment of AI Models
- Online-testing Fraud Prevention System for Detecting Spoofing
- Spherical Drone Display

Toward the Co-creation of New AI Services

General Manager of Service Innovation Department

Masayuki Tsuda, Ph.D.

Artificial intelligence (AI) is booming again, and hardly a day goes by without some mention of AI in newspapers and on news sites. As in previous AI booms, there are some overexpectations and misunderstandings about AI. However, emerging AI services using image recognition, voice recognition, and other technologies are raising expectations that AI will actually be more practical this time.

Shortly before the previous AI boom (which was from the second half of the 1980s to the first half of the 1990s) came to an end, the neural network*1 attracted much attention as a promising learning approach. At that time, I was a student doing research on controlling a robot using a neural network with a back-propagation*2 algorithm. In my experiments, a mainframe computer spent the night processing the data collected by sensors and learning motions from the data. The results of learning were downloaded to a notebook computer and used to control the robot. At that time, many people including me greatly expected neural networks to provide the breakthrough needed to make a computer intelligent enough to enable innovative services such as autonomous driving. However, the then neural networks didn't have such capabilities, so the expectations and boom withered away before we knew it. Well, about 25 years have passed since then, and a new AI boom has been sparked by deep learning using multi-level neural networks. Technology development seems to run in cycles in much the same way as business.

What are the differences between the current AI boom and the previous ones? One difference is that the learning algorithms have improved; the recurrent neural networks in use today can more easily learn from time-series data, and convolutional neural networks can more easily learn from spatial data and have found a wide range of application. This is the result of steady and patient efforts by AI researchers over the years. The other differences relate to the spread of the Internet and major changes in the computing environment that have contributed to the resurgence of AI. The new learning algorithms can now be released as open source software over the Internet and updated as needed. Cloud computing, meanwhile, has made it possible to process large volumes of diverse data (big data) in realistic periods of time. As a result, even software developers without any specialized skills in AI can now develop in a relatively short time AI services utilizing terabyte-class big data and cutting-edge deep-learning programs released over the Internet.

Hurdles to development of AI-based services have therefore been lowered drastically in comparison with the past, and we can expect competition in the development of such services to accelerate. In such an environment, how can one AI-based service differentiate itself from others? My answer is "big data." With deep learning, for example, one approach is to increase the types of data for smooth convergence of learning errors and to use a huge amount of data for improved learning accuracy. Consequently, to achieve even faster development of more valuable services, how to acquire big data suitable for a service is a key issue.

A useful approach to resolving this issue is "co-creation (+d*3)" promoted by NTT DOCOMO. One example of this approach is the "AI Taxi" service, which predicts customer demand for taxis. This novel service was developed through co-creation with a taxi company. AI taxi consists of two parts. One part creates a prediction model using deep learning and a combination of big data possessed by NTT DOCOMO and big data possessed by its partner taxi company. The other part processes real-time information using that model and predicts the areas in which customers will be looking for a taxi 30 minutes later. The prediction is done every 10 minutes. Experiments have shown that the service can increase the income of drivers, so the taxi company has decided to use the service in its business. Needless to say, AI Taxi could not have been developed with only one or the other type of big data. It came about, rather, through co-creation activities that combined the big data contributed by each company. The co-creation of a new service or new business by combining big data in this way can also be accomplished between NTT DOCOMO and other partners. A new service born in this way could be used for the business of a party other than NTT DOCOMO or its partner. From the viewpoint of utilizing big data and AI, co-creation will become more important in the near future.

Going forward, through co-creation activities, my department and I commit ourselves to developing new services and new businesses such as AI Taxi by combining our own big data with our partner's big data and by applying the power of AI.

---

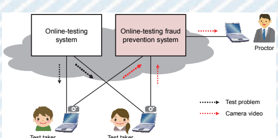
*1 Neural network: An information processing system that simulates the brain's nerve cells (neurons) and their interconnections. Learning is achieved by changing the strengths of those connections.

*2 Backpropagation: A learning algorithm used by neural networks for changing the strengths of inter-neuron connections by propagating neuron output errors backwards to the neurons to minimize error.

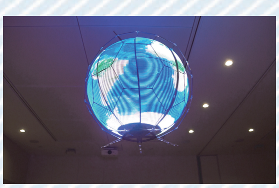
*3 +d: Brand name of NTT DOCOMO initiative for creating new value together with partner companies.

NTT DOCOMO
Technical Journal
Vol.19 No.4 Apr. 2018

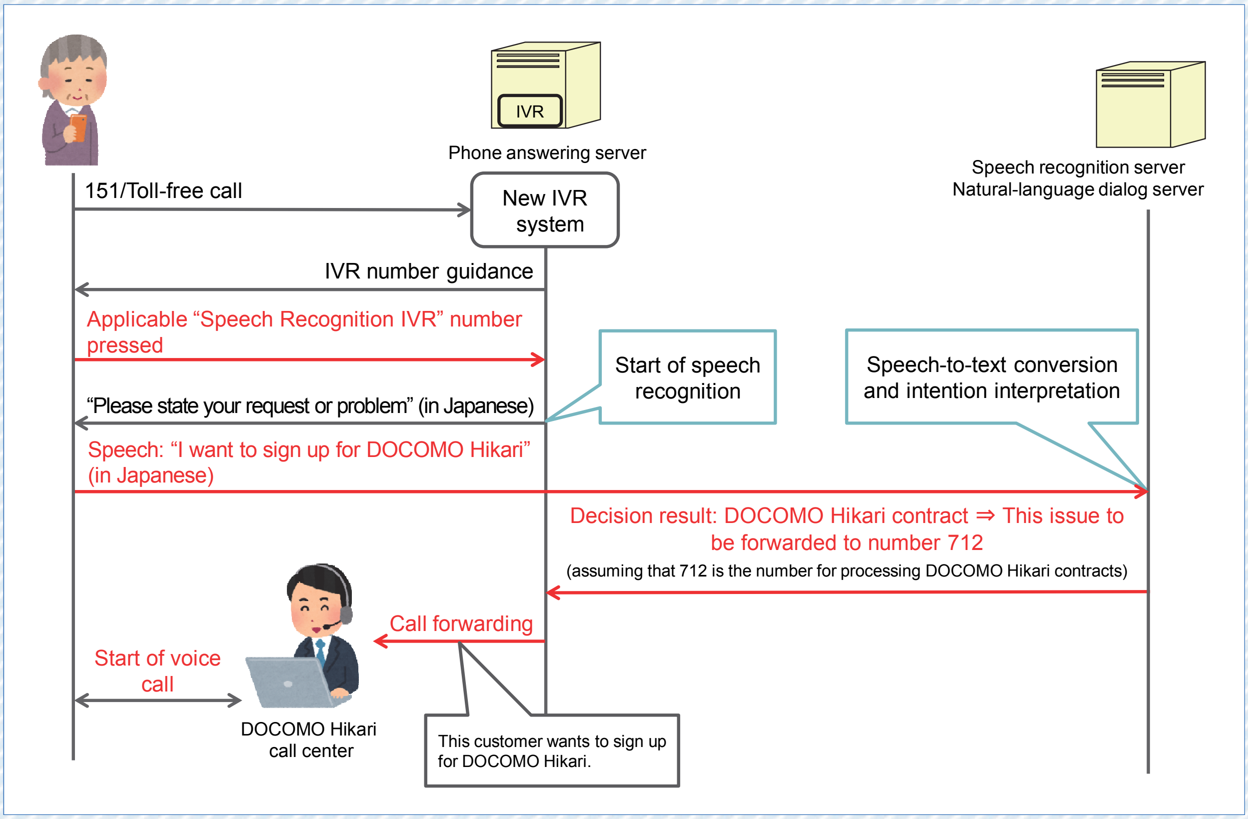
[Contents]



(P.19)



(P.26)



Technology Reports Improving Customer Satisfaction and Operator Efficiency in Call Centers Using AI —Speech Recognition IVR— (P.4)

Overview of Speech Recognition IVR service

Technology Reports

Call Centers / Speech Recognition / AI

Improving Customer Satisfaction and Operator Efficiency in Call Centers Using AI

—Speech Recognition IVR—

Service Innovation Department Takanori Hashimoto Yuuki Saitou Yuriko Ozaki

With recent advancements in AI technology, customer interaction systems using speech have started to be implemented. NTT DOCOMO has developed a Speech Recognition IVR system that can be introduced to receive telephone inquiries, and uses speech recognition to forward calls to the appropriate specialized call center. This improves customer satisfaction by reducing wait times and eliminating the need to enter numbers, and improves operator efficiency. This article describes the structure of Speech Recognition IVR and how it came to be introduced.

1. Introduction

The speech guidance for NTT DOCOMO's general inquiries line, called Interactive Voice Response (IVR)*1, has begun providing a Speech Recognition IVR functionality that is able to forward calls to the most suitable number by simply having the caller say what they need. This service can be used by dialing 151 from a DOCOMO mobile phone, or by calling the Free-Dial®*2 number (0120-800-000) and selecting the Speech Recognition IVR number.

Previously, customer inquiries were handled through the main, general inquiries call center together with various other specialized call centers such as the 113 Center (for repairs) and the DOCOMO Hikari call center. The basic process was for customers to call the general inquiries number and be forwarded to the appropriate center by selecting a number based on the voice guidance, but because there were many numbers to choose from and it was often difficult to know which one was the best, many customers just asked the operator to forward them to the appropriate center. In fact, of all the calls handled by the general call center,

©2018 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

*1 IVR: Automatic voice response equipment. A system that provides guidance on the telephone, with phrases such as, "For help with …, please press number…."

*2 Free-dial®: A registered trademark of NTT Communications Corp.

approximately 20% of the customers were not able to select the correct option after voice guidance for reasons like those stated above. These callers needed to be forwarded from the selected specialized center to another one. As a result, operators needed to handle call forwarding and other inquiries for these callers. This created more congestion, increased customer wait times, and occupied specialist operators understanding customer inquiries a second time after the general reception center had forwarded the calls to them.

Speech interaction services in various forms have been implemented and offered in recent years. NTT DOCOMO has provided the "Shabette-Concier" service, which infers the intention of user utterances and performs searches for weather and restaurant information, and can launch the telephone applet.

NTT DOCOMO also provides the "DOCOMO DriveNet" [1] application, which provides useful information to drivers by simply talking to it, and has developed OHaNAS®*3 in collaboration with TOMY Co. Ltd., as part of NTT DOCOMO's "+d"*4 initiative to work with partners to create new value. OHaNAS is a toy that operates through a smartphone or tablet and is able to have natural conversations. NTT DOCOMO is developing a natural-language dialogue platform, as described above [2], and is accumulating knowhow in speech interaction services. As such, we have introduced Speech Recognition IVR utilizing AI technology, to improve issues with speech guidance systems as described earlier.

The system enables customers to simply state their issue, without needing to do complex button operations, to be connected to the appropriate call center. This reduces customer wait times before reaching an operator that can handle their issue, compared with forwarding from the general inquiries center. It also reduces work transferring calls by operators at the general inquiries center, and the customer's inquiry is converted to text and can be sent to the operator it is forwarded to, which can help reduce time in initial handling of the call.

This article describes the structure of the Speech Recognition IVR system.

2. Overview of Speech Recognition IVR

Speech Recognition IVR is a system combining IVR, which can receive inquiries by telephone, and speech recognition/intention interpretation*5, and is able to transfer calls to the appropriate specialized call center. The basic process of using the system is shown in **Figure 1**. After the IVR guidance is played, the user selects the number for "Speech Recognition IVR," and speech recognition begins. The speech recognition function converts the customer's statement of their issue to text. Then, the intention interpretation function determines which operator to forward the call to among all of the specialized call centers, who has skills appropriate for the issue. The telephone reception server receives the above result from speech recognition and intention interpretation, and forwards the call to the appropriate center. The text version of the customer's issue is also sent to the operator.

The above system connects the customer to the appropriate center, when they simply select the

*3 OHaNAS®: A registered trademark of TOMY Company, Ltd.

*4 +d: Name of NTT DOCOMO initiative for creating new value together with partner companies.

*5 Intention interpretation: A function that interprets the intention behind the text and assigns it to an appropriate task. An example of text and a task is, "Will it rain today?" and "Weather report search."

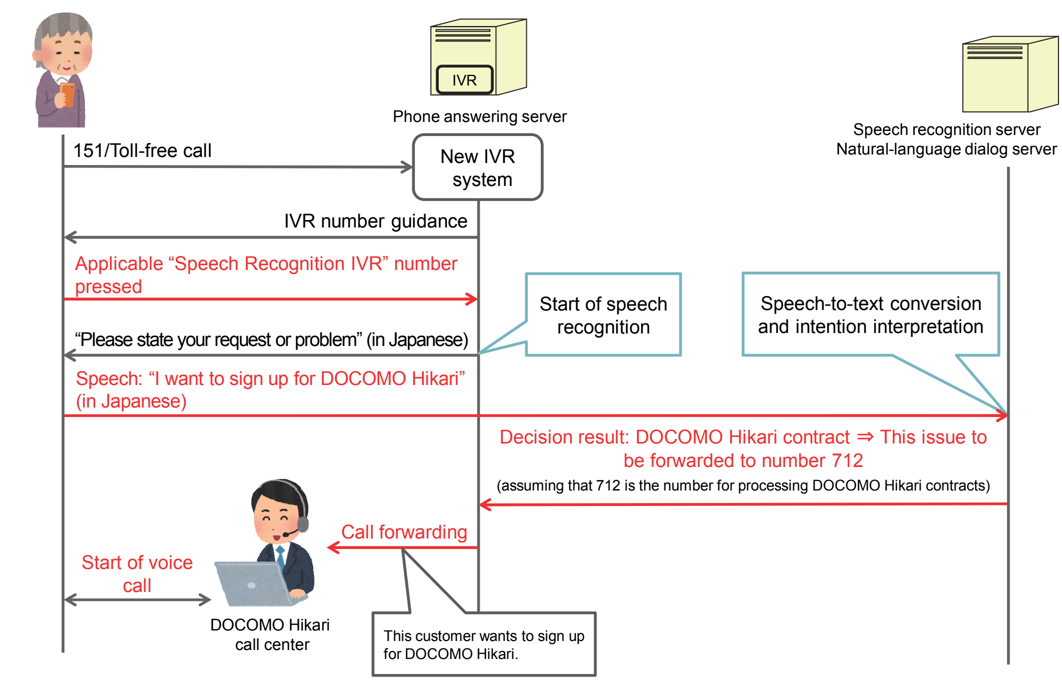


Figure 1　Overview of Speech Recognition IVR service

Speech Recognition IVR number and stating the issue. The operator is also able to check a summary of the issue before speaking with the customer.

3. Speech Recognition Function

The speech recognition function converts the recorded speech of the calling customer to text. This process is shown in **Figure 2**. Conversion to text uses an acoustic model*6, a pronunciation dictionary, and a language model*7. The speech recognition technology first uses the acoustic model to match phonemes*8 to the utterance waveform, and then converts the phonemes to words using the pronunciation dictionary. The language model is then used to statistically determine the word order, which is finally converted to text in sentences. Features of each of these steps are described below.

(1) Customers calling the call center vary in aspects such as gender, age, and location, so the sound of the utterance also varies. Calls also come from different phones, whether fixed-line or smartphone, and each use different types of audio compression, so the sound quality varies. The acoustic model

*6 Acoustic model: Statistical model comprising frequency characteristics of phonemes targeted for recognition possesses.

*7 Language model: Statistical model comprising morpheme arrangements and frequency of those arrangements.

*8 Phoneme: The smallest units of speech, such as vowels and consonants.

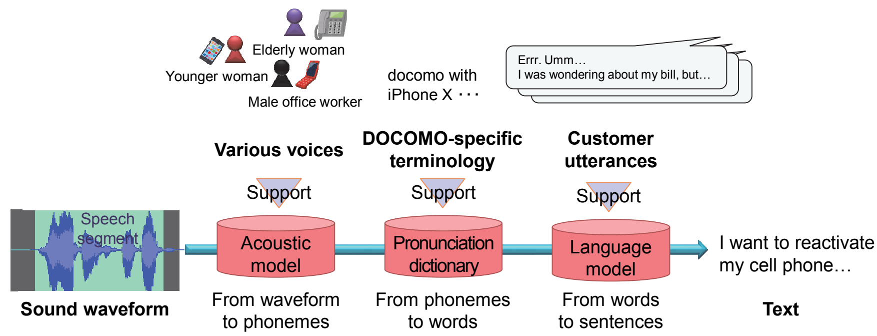


Figure 2　Speech recognition system processing

handles such differences in sound quality.

(2) Words that can be expected to occur in a telephone inquiry and related to NTT DOCOMO (handset models, service names, etc.) together with variations in pronunciation are registered in the pronunciation dictionary and used to improve word recognition accuracy.

(3) The customer utterance is not a well-formed, carefully recited sentence, but is filled with fillers*9 and other modified phraseology, much like free conversation. To analyze such sentences, as many as possible anticipated phrases of this type are gathered and used to train the language model.

Accurate and appropriate ending points in the customer's utterance are also found, and it is forwarded without unnecessary sections to further improve usability. In doing so, there is also a function to segment utterances naturally, taking noise (ambient, other speakers, music, etc.) in the customer's environment into consideration.

The calling customer's statements are converted to text using these speech recognition functions and passed to the intention interpretation function.

4. Intention interpretation Function

1) Overview

The intention interpretation function determines what sort of inquiry is intended in the utterance, from the customer text information, and decides where the call should be forwarded. Destinations for forwarding calls are decided based on the numbers provided by NTT DOCOMO telephone inquiry reception, as shown in **Table 1** (as of September 2017). Our customer inquiries span an extremely wide range, so to increase coverage, we have added items for inquiries not included in Table 1.

Interpreting the intention of the customer's utterance can be solved as a sentence classification problem, classifying the input text. Specifically, a

*9 Fillers: Words used to connect utterances, like "umm…" and "you know…."

Table 1 DOCOMO inquiry categorization

Major category	Minor category	Sub-category
Inquiries regarding DOCOMO Hikari	Moving, new applications and other procedures and inquiries	Moving procedure (transfer)
		New subscriber procedure
		Cancellation procedure
		Inquiry
	Construction date inquiries	Construction scheduling
		Construction rescheduling
		Other construction inquiries
	Connecting and settings	
	Problems	
Suspending/ Resuming service		
Operation of telephones and data communications devices	iPhone and iPad® operation	
	Smartphone and tablet operation	
	docomo Feature Phone operation, settings	
	DOCOMO Hikari Internet and other settings	
	Mobile Wi-Fi® router and other data communication device operation and settings	
Problems/ Service area	Smartphone and tablet problems	
	docomo Feature Phone and other terminal problems	
	Service areas	
	iPhone and iPad® problems	
	DOCOMO Hikari problems	
Guidance on various services	Rate plans, discount services	
	Mobile number portability	
	Various network services	
	International services	
	Unlocking a SIM lock	
Other order procedures/ inquiries	Order procedures	
	Inquiries	

iPad®: A trademark of Apple Inc. registered in the USA and other countries.

Wi-Fi®: A registered trademark of the Wi-Fi Alliance.

model is trained using the language dictionary and a large number of sample utterances matched to forwarding numbers, as shown in **Figure 3**. It is able to determine the intention and a forwarding number from the input text at high speed. As with tuning of speech recognition, the sample utterances and language dictionary are created using NTT DOCOMO specific terms and covers the diversity of utterances occurring in inquiries.

2) Discriminating Ambiguous Statements

There are cases when it is difficult to determine a suitable forwarding number from the limited information in a customer's statements. For example, for an inquiry like, "How do I use my device?" although it is clear that it is about using a device, there is a wide range of devices handled

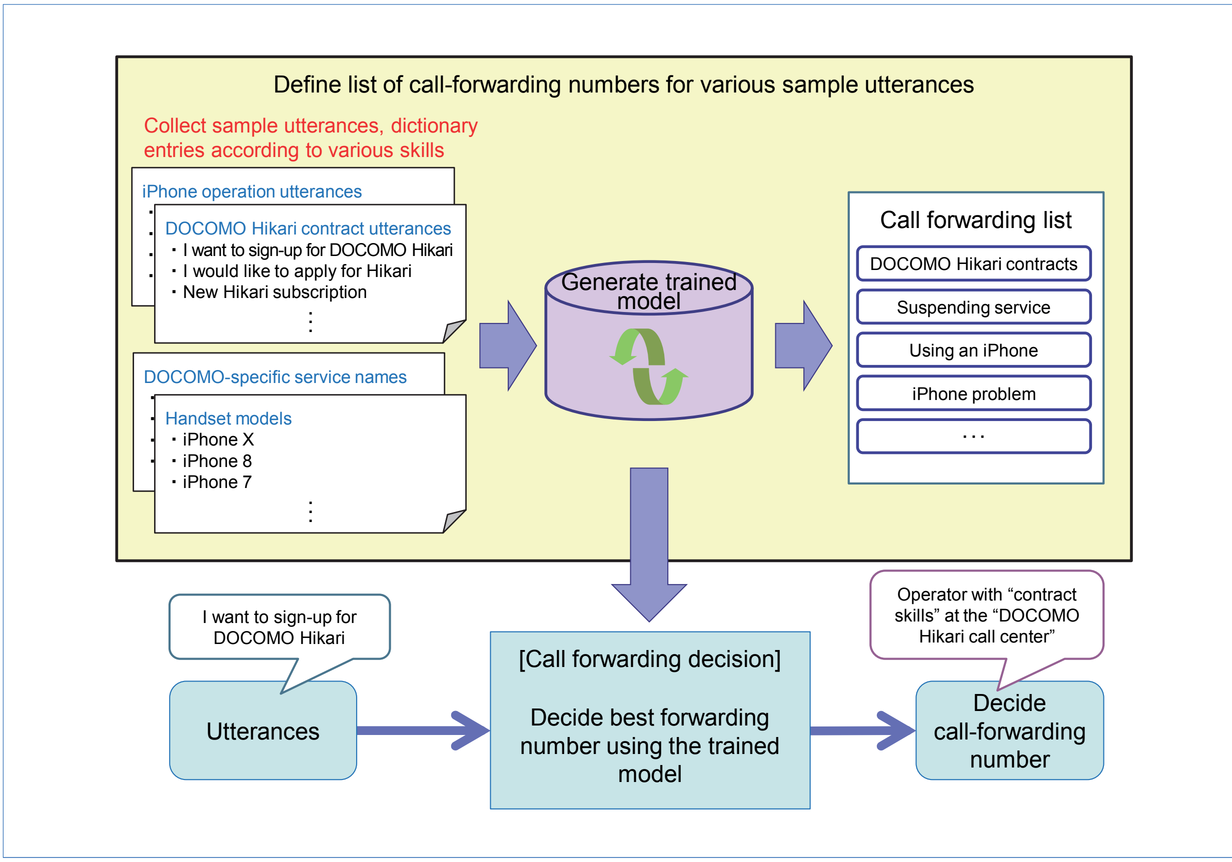


Figure 3 Deciding call forwarding number

by NTT DOCOMO and it can be difficult to narrow this down. As such, we use two methods to discriminate ambiguous utterances.

- The first is to prompt the user. For example, if it is about how to use a device, we can narrow down to specialized centers by asking about major categories (iPhone®*10, Android™*11, docomo Feature Phone, etc.).
- The second is to narrow-down and forward the call according to operator-specific skills. Within each specialized center, each operator is able to handle a different range of inquiries. Inquiries that are more ambiguous can be forwarded to operators with broader skills, and if the utterance contains enough information, it is forwarded to an operator with suitable skills.

These two approaches are used to forward inquiries to an appropriate specialized center according to the content of the customer's utterance.

The utterance and forwarding number obtained using the speech recognition and intention interpretation processes described above are used by the telephone reception server to decide where the customer's call will be forwarded and are displayed on the operator's screen.

*10 iPhone: A registered trademark of Apple, Inc. United States, used within Japan under a license from Aiphone Co., Ltd.
*11 Android™: A trademark or registered trademark of Google Inc., United States.

5. Conclusion

This article has described a Speech Recognition IVR system that introduces AI into call center work and is able to transfer a call to an appropriate specialized call center and send details to the operator by just having the caller speak on the telephone. After the system was introduced, 10 to 20% of customers were transferred to the specialist center, and this was the goal we initially intended, without using the general reception center, reducing the work of operators transferring calls. We will continue efforts to improve performance, increasing use of Speech Recognition IVR and the rate of successful call forwarding.

This article has only dealt with forwarding of calls, but we are studying ways to automate parts of call center work, by expanding the interaction of customers with the AI to be bi-directional. We are also studying the potential to provide this functionality to other companies as an NTT DOCOMO corporate solution.

REFERENCES

[1] R. Kurita et al.: "DOCOMO DriveNet Info," NTT DOCOMO Technical Journal, Vol.16, No.1, pp.4–10, Jul. 2014.

[2] K. Onishi et al.: "Natural-language Dialogue Platform for Development of Voice-interactive Service," NTT DOCOMO Technical Journal, Vol.17, No.3, pp. 4–12, Jan. 2016.

Technology Reports — Deep Learning | Container Virtualization | Parallel Processing

Deep Learning Platform Capable of Rapid Creation and Deployment of AI Models

Service Innovation Department Takero Ibuki†

AI development using deep learning is being conducted enthusiastically in recent years, but to realize an accurate AI, developers must run multiple learning trials, modifying learning parameters on a trial-and-error basis. This trial-and-error process is a major factor in increasing development time. There are also many frameworks and libraries for performing deep learning, and it is becoming a complex task just to manage and update these systems. NTT DOCOMO has developed a deep learning platform to resolve these issues, enabling highly accurate learning in a short period of time. This article gives details of this platform and describes the technology.

1. Introduction

Deep learning using deep neural networks has attracted much attention recently in Artificial Intelligence (AI), artificially performing tasks such as image recognition, translation, and speech recognition using computers. Neural network*1 system technologies have existed for some time, but due to issues with accuracy, they were not practical. However, major advances in General Purpose computing on Graphics Processing Unit (GPGPU)*2 computing resources, Compute Unified Device Architecture (CUDA)®*3, and the development of the NVIDIA CUDA Deep Neural Network library (CuDNN)*4, have made it possible to implement deeper neural networks, greatly increasing their accuracy. Many frameworks*5 for developing deep neural networks, such as Tensorflow®*6 and Caffe*7 have appeared, and AI development using deep learning is now being done in many research facilities and enterprises. NTT DOCOMO is also using deep learning technology in AI services such as

©2018 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

† Currently DOCOMO Innovations, Inc.

*1 Neural network: An entity that numerically models nerve cells within the human brain (neurons) and the connections between them. It is composed of an input layer, an output layer and intermediate layers and are able to approximate complex functions by varying the number of neurons and layers and the strength of connections between layers.

the image recognition Application Programming Interfaces (API)*8 being provided by docomo Developer Support. In development of this AI, multiple learning trials under different conditions needed to be done in order to obtain accurate results, and since these were done manually, it required much manpower and did not utilize servers efficiently. It was also difficult to build and update the multiple libraries*9 and framework environments used.

For these reasons, NTT DOCOMO has developed a deep learning platform using container virtualization technology*10 and an original scheduling system to process large volumes of conditions automatically. This has enabled highly accurate learning in short periods of time.

This article describes details of the technologies used in the deep learning platform developed by NTT DOCOMO, and the effects brought about by the platform.

2. Deep Learning and Related Issues

2.1 Overview of Deep Learning

Deep learning is a type of machine learning*11 for technologies using deep neural networks, which are neural networks having many intermediate layers.

Deep learning consists of the two phases: the learning phase, in which patterns and rules are learned from large amounts of data, and the inference phase in which the result of learning is used to infer which patterns apply to the new data. The neural network after learning is referred to as an AI model.

There are a wide range of AI models and services, such as image recognition using image data, or speech recognition and translation using speech data.

2.2 Tuning Parameters

For deep learning, it is not the case that by simply preparing the data, the best patterns and rules will be learned automatically. Learning conditions such as initial values and learning rates, which are called hyperparameters, must also be configured [1] [2]. Some typical hyperparameters are shown in **Table 1**.

Even with the same data, different hyperparameters

Table 1 Typical hyperparameters

Hyperparameter	Description
Minibatch size	Number of data items processed in a single trial
Learning rate	Breadth of weighting updates
Number of iterations	Number of trials
Weighting initial value	Initial value of neuron connection strength
Activation function	Neuron activation method
Optimization function	Method for deciding how learning progresses
Drop-out rate	The rate for disabling neurons in a given node

*2 GPGPU: The use of GPUs, which are generally used in computers for rendering and other types of image processing, for other types of applications. GPGPU excels at parallel distributed processing.

*3 CUDA®: A general-purpose parallel programming environment for GPUs provided by NVIDIA. A registered trademark of NVIDIA Corp.

*4 CuDNN: A deep learning CUDA library (See *8) published by NVIDIA. A registered trademark of NVIDIA Corp.

*5 Framework: Software that encompasses functionality and control structures generally required for software in a given domain. With a library, the developer calls individual functions, but with a framework, it handles the overall control and calls individual functions added by the developer.

*6 Tensorflow®: A deep learning programming framework (See *5). A registered trademark of Google Inc.

*7 Caffe: A deep learning programming framework (See *5).

can result in issues such as no improvement through training, or a problem called over-learning, in which the network loses generality and becomes biased to specific data only. Selecting the best hyperparameters is extremely important to building a deep learning model.

There are many types of hyperparameter, and setting them manually by trial-and-error requires large amounts of human work. Manually running trials while changing hyperparameters requires logging into each server manually and checking the server state before each trial, which results in poor server utilization and increases the time required to obtain results.

2.3 Building Environments

Both the learning and inference phases of deep learning are composed of many matrix operations. For this reason, GPUs, which are capable of parallel processing and have more cores than a CPU, are able to accelerate deep learning. GPUs from NVIDIA are widely used. GPUs were originally designed for image processing, but they are also used for high-speed computing for deep learning and other applications, and there are programming environments and libraries such as CUDA, for general processing, and CuDNN for deep learning. Because using these programming environments and libraries requires learning specialized languages that extend the C programming language, which takes time, universities and enterprises have developed frameworks that make deep learning programming easier and these have been released as Open Source Software (OSS)*11. Typical frameworks include Tensorflow and Caffe. Advantages of these frameworks are that they can be used with Python*12, which is often used for data analysis, and there are API that make it easier to code neural network structures.

However, these libraries and frameworks must be kept up-to-date to ensure that they will run and produce the fastest processing speeds. Keeping up with all updates is labor intensive and can be quite complex due to dependency issues such as library versions that only support certain GPUs and frameworks only supporting certain versions of libraries.

3. Deep Learning Platform

3.1 Approaches to Resolving Issues

Below we describe methods for resolving the two issues described above: (1) inefficient human work in selecting hyperparameters, and (2) complexity in building environments; and the deep learning platform system that we have developed.

1) A System to Process Learning Jobs in Parallel, Automatically According to Server State

The following observations apply to issue (1) above.

- Parameter changes and learning processes are performed manually and successively.
- The state of each server is monitored manually.

Our approach to resolving these problems was to control a job queue and execute jobs automatically in parallel, according to server state. We developed a system able to manage servers centrally and run a sequence of learning processes in

---

*8 API: A set of rules that define programming procedures used to access commands and functions when software is developed for a particular platform (OS or middleware).

*9 Library: A collection of high-versatility programs in a reusable form.

*10 Container virtualization technology: A type of computer virtualization technology in which dedicated areas called containers are created on the host OS and the application software runs within these containers.

*11 Machine learning: A framework that enables a computer to learn useful judgment standards through statistical processing from sample data.

parallel.

2) Adoption of Container Virtualization Using Docker®*14

The following observations apply to issue (2) above.

- Updating libraries and frameworks consumes time.
- Dependency relationships between libraries and frameworks make updates difficult.

As a solution to these issues, we focused on a virtualization technology that enables us to build environments independently for each job within the server. In this case, we only needed libraries and frameworks for deep learning and not OS functionality as in conventional host virtualization technology*15, so we decided to use container virtualization technology, and specifically, Docker (**Figure 1**). Docker is software that uses Linux®*16 container virtualization technology, which isolates resources used by the OS so that multiple environments can be established. It is faster than conventional virtualization technologies such as hypervisor virtualization*17 and host virtualization because it does not require a new OS to be started. It also has the benefit that once a container image has been created, it can be moved to another environment easily.

3.2 System Overview

A system overview of the deep learning platform we have developed is shown below (**Figure 2**). The platform is composed of a Master server, Worker

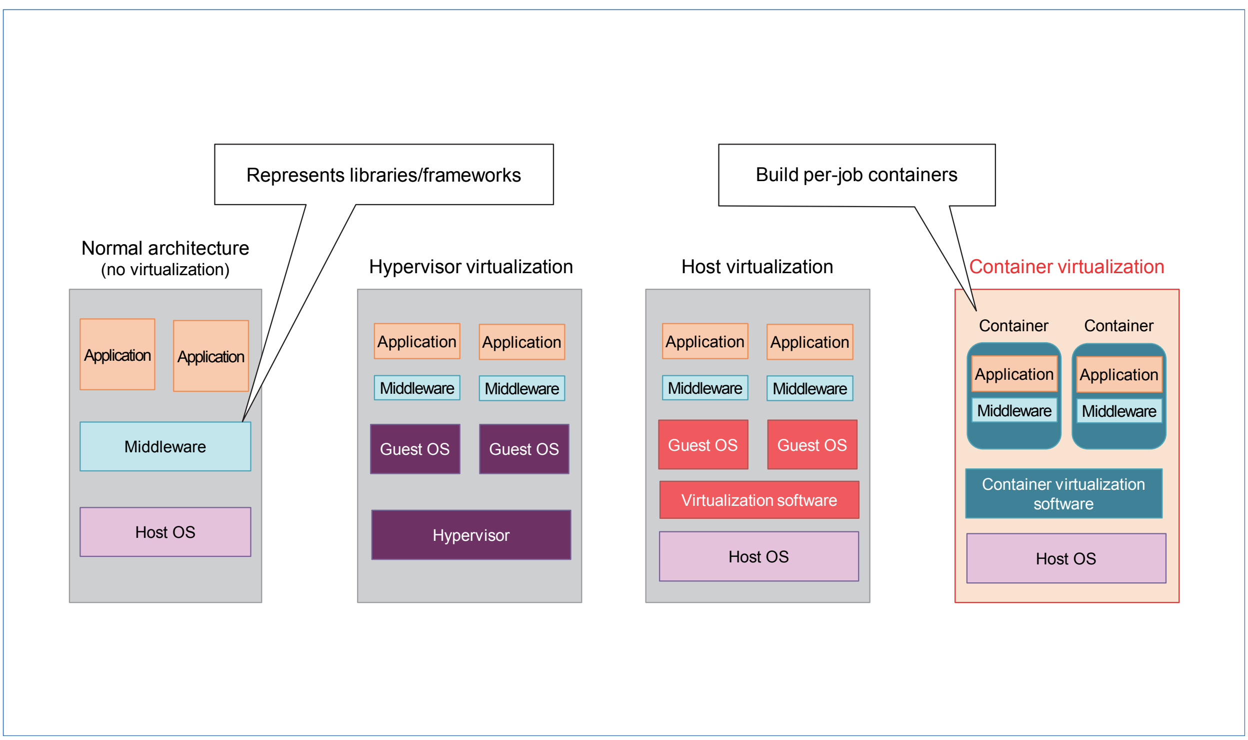


Figure 1 Differences between container virtualization and other virtualization technologies

*12 OSS: Software for which the source code is published free-of-charge, and anyone can re-use or modify it. However, the original author retains the copyright, and even if a derived work is created or it is redistributed, expression of the original author's copyright must be maintained.

*13 Python: A general purpose programming language.

*14 Docker®: Container-type virtualization software. A registered trademark of Docker Inc.

*15 Host virtualization technology: A type of computer virtualization technology. Software that acts as a base for the OS is installed on the server, and virtual machines run on that software.

*16 Linux®: An open-source Unix-type OS that can be freely redistributed under GNU Public License (GPL). A registered trademark or trademark of Linus Torvalds in the United States and other countries.

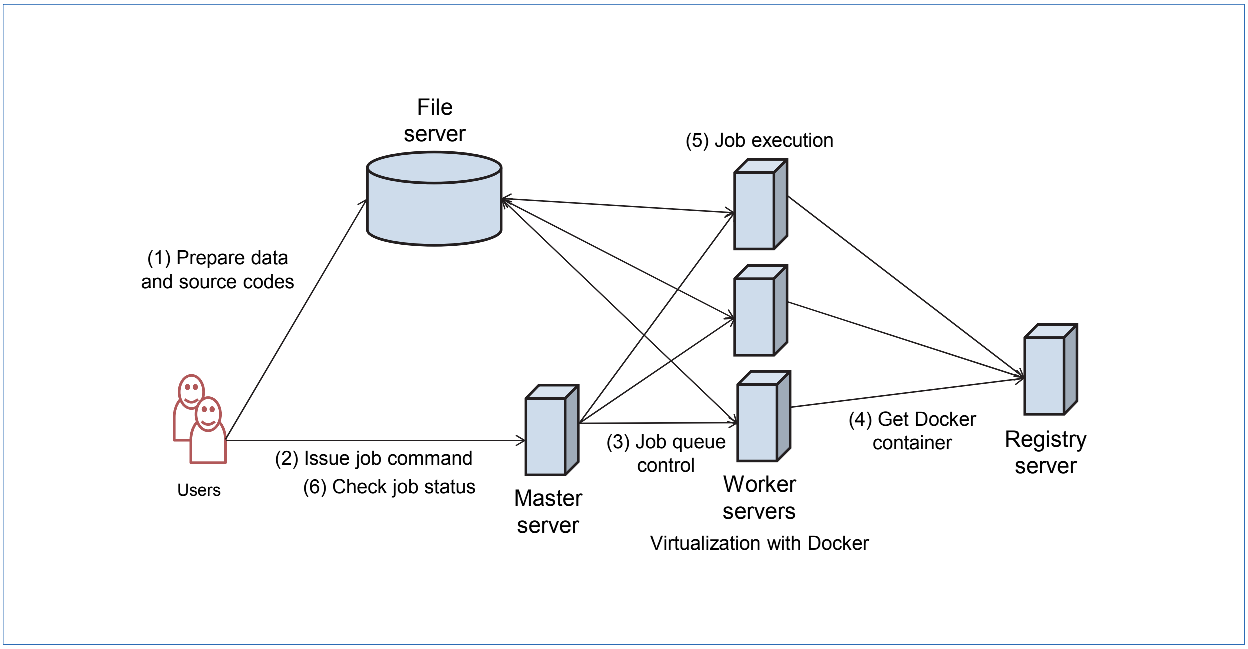


Figure 2　Deep learning platform system architecture

servers, a File server, and a Registry server.

- The Master server manages all deep learning jobs, and all input from users is processed through the Master server.
- Worker servers receive instructions from the Master server and perform learning tasks. Learning tasks are performed in Docker containers in each of the Worker servers. Note that Docker alone does not include functionality to use the GPU on a Worker server, so the Nvidia-docker*18 OSS is used.
- The File server is used to store deep learning data and source code.
- The Registry server is used to store the Docker Registry, which centrally manages the Docker containers. Docker containers with many combinations of frameworks and libraries can be created and stored in the Registry server so that using the platform, users can select any of the deep learning framework and library combinations they need.

Users perform two tasks when using the system: storing deep learning data and source code, and registering a job in the Master server. Job control and execution is performed automatically.

3.3 Job Management

The following functions were developed for the Master server, which manages jobs submitted by users.

(a) Receiving and registering jobs (Fig. 2 (2))

(b) Controlling the job queue (Fig. 2 (3))

(c) Building run-time environments for jobs and executing jobs (Fig. 2 (3)–(5))

*17 Hypervisor virtualization: A type of computer virtualization technology. Virtual machines run installed directly on the server.

*18 Nvidia-docker: Docker extension software.

(d) Monitoring job state (Fig. 2 (6))

We studied two ways to implement the job registration and job control functions: using a message queue*19, and using a Relational DataBase (RDB)*20. Using a message queue yields greater throughput and lower latency than a RDB, but monitoring of job states and processing retries is more difficult to implement. We adopted the RDB method in this case, because job monitoring was important, and since actual job execution is done by Worker servers, throughput and latency of job registration and queue control was less important.

Next we describe specifically how these functions were implemented.

The user prepares the data and source code required for the deep learning task (Fig. 2 (1)). Then, a job command in a pre-determined format is registered on the Master server (Fig. 2 (2)), and a queue entry in the form of an RDB record is created and added in the Master server. Job commands and options were developed for this platform so that the deep learning framework can be selected, and the location of data and source code on the File server can be specified.

For job queue control, it was desirable to only run a single job on one GPU due to GPU memory restrictions. As such, job control was implemented by having the Master server periodically check the status of GPU usage on each Worker server (Fig. 2 (3)), and run the next job on the queue if it is idle (Fig. 2 (3)). For jobs that are ready to run, the allocated Worker server retrieves the specified Docker container from the Registry server (Fig. 2 (4)) and runs the job (Fig. 2 (5)).

The last function is job state monitoring, which the user can do by referring to the RDB table (Fig. 2 (6)). Note that, as mentioned earlier, (b) and (c) are controlled and performed automatically.

4. Results

There are three main results achieved through introduction of this platform.

1) Reduced AI Model Training Time

Previously, when developing AI models, each run and set of learning conditions were run together on a server, as shown in **Figure 3** (a). This resulted in low server utilization and much waste. This platform enables job scheduling to be automated and run in parallel, so that significant reductions in training time are achieved (Fig. 3 (b)).

2) Improved Learning Accuracy

With an increase in efficiency of AI model training, training can be attempted under more conditions, which has achieved great increases in training accuracy (**Figure 4**).

3) Accelerated Environment Building

Use of Docker has made it easy to introduce new deep learning frameworks or updates. In concrete terms, the several hours it previously took to build an environment has been reduced by a factor of ten.

5. Future Prospects

Three items present themselves as future prospects.

*19 Message queue: A mechanism used when exchanging data and cooperating between application software, in which data to be transmitted is temporarily stored without waiting for the other application to process it, and the sending application continues on with its next task.

*20 RDB: A data base management scheme in which data is managed in multiple tables and complex data relationships are handled by defining relationships among tables.

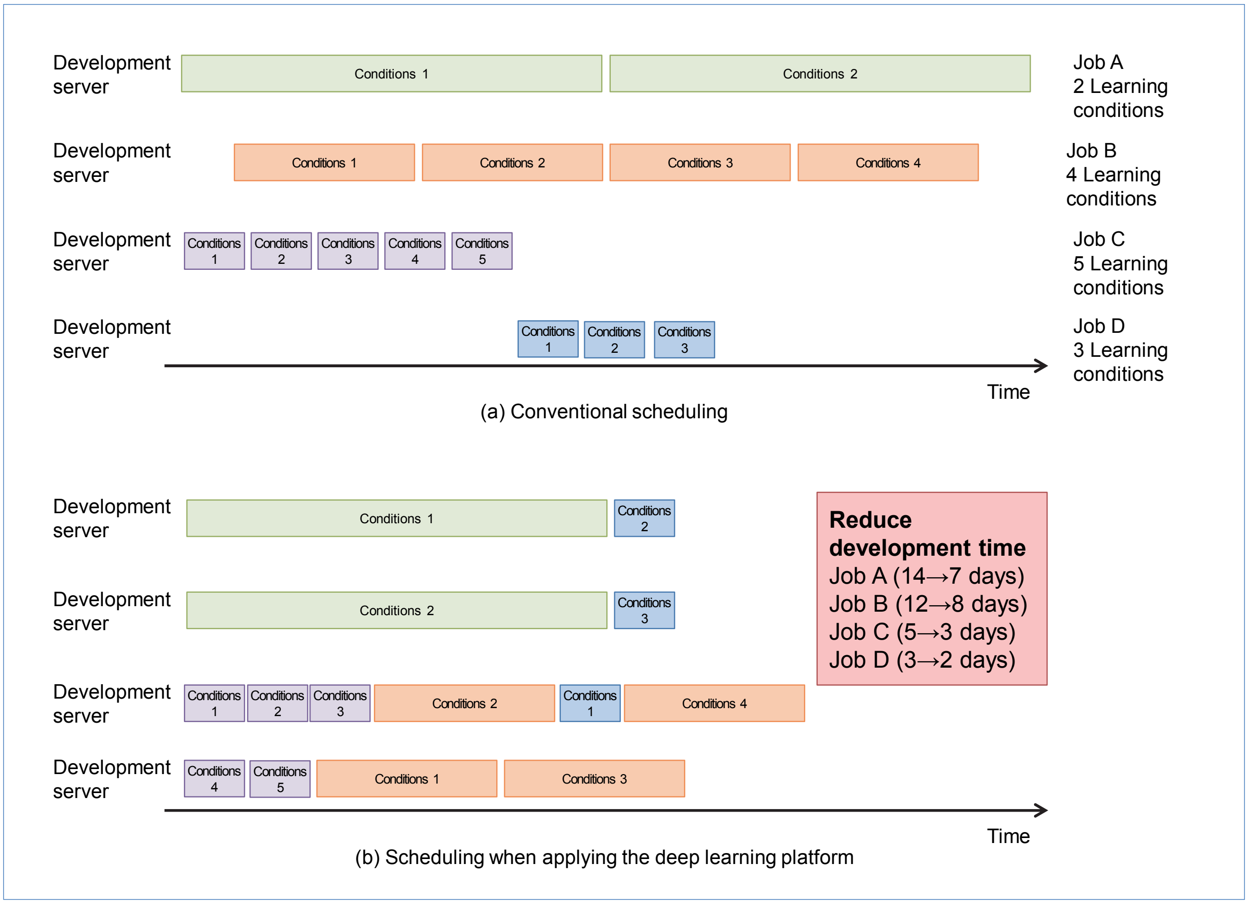


Figure 3　Effects of applying the deep learning platform

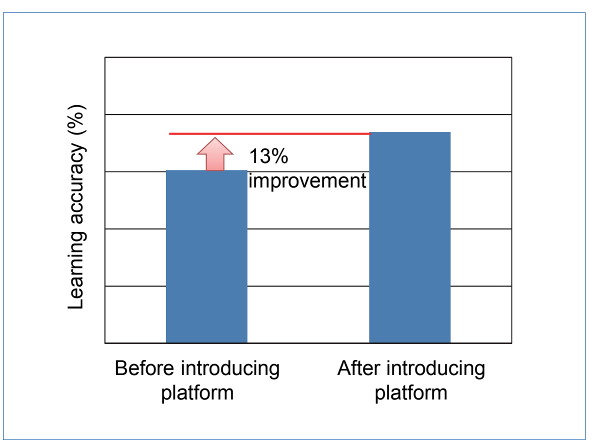


Figure 4　Improvement in accuracy through use of the deep learning platform

1) Support multiple generations of GPU, multiple cloud environments

Generations of GPU advance almost every year, and the size of model that they can handle also changes, so jobs must be distributed to Worker servers with consideration of the GPU generation. Implementation must also be consolidated to handle multiple cloud environments.

2) Upgrade the Scheduler

The current job scheduler can only apply a simple priority to jobs, so this will be improved to allow application of more complex priorities.

3) Optimize Parameters

Current training only allows multiple conditions to be run in parallel, so the more conditions are submitted, the more the servers will be monopolized. By establishing algorithms for selecting optimal combinations of parameters that can efficiently reduce computing time, it will be possible to develop accurate AI models in less time.

6. Conclusion

In this article, we have described technical details of a platform that facilitates deep learning processes for developing accurate AI models in short periods of time. We intend to use this platform to accelerate deep learning initiatives, to advance the platform itself, and to improve efficiencies.

REFERENCES

[1] T. Okatani: "Deep Learning," Kodansha, 2015.
[2] Y. Bengio: "Practical recommendations for gradient-based training of deep architectures," Neural Networks, Springer, pp.437–478, 2012.

Technology Reports
Face Authentication / Online Testing / Fraud Prevention

Online-testing Fraud Prevention System for Detecting Spoofing

Service Innovation Department Kazunori Yamamoto Kazuhiko Ishii Kimihiko Sekino

It is becoming increasingly important to prevent fraudulent activity such as spoofing (impersonation) in online testing that allows individuals to take tests at home and other remote locations. One method of preventing such activity is to have the test proctor monitor the student remotely by camera, but this presents issues from the viewpoint of work efficiency. We have developed an online-testing fraud prevention system that supports test proctors by applying face authentication technology to automatically detect spoofing during an online test. This system can make the proctoring of online tests more efficient.

1. Introduction

The increasing popularity of distance learning [1]–[3] in recent years has led to a desire for online testing that allows students to take tests over the Internet regardless of time or location. In online testing, however, no test proctor is present, so opportunities for inappropriate behavior exist. In addition, fraudulent activity that would be difficult to imagine at a traditional testing center can occur, such as having someone else take the test or receiving answers from someone nearby. The prevention of fraudulent activity by test takers is just as important in online testing as it is in tests taken at traditional sites, and services responding to this need have begun to be provided [4]–[7].

Fraudulent activity at the time of a test includes someone other than the student taking the test (spoofing) and obtaining information during the test by some means (cheating). Examples of

©2018 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

countermeasures to such activity in existing online-testing fraud prevention services are listed in **Table 1**. Although existing services have automated some procedures such as biometric authentication for identity verification at the beginning of an online test, they tend to rely on manual methods for preventing fraud during a test by having a proctor remotely monitor the test taker by camera. It therefore stands to reason that automatic detection of fraudulent activity during a test could make test proctoring much more efficient.

In this article, we focus on the automatic detection of spoofing during a test through the application of face authentication technology. We take up the automatic detection of cheating in future research. Preventing spoofing in online testing requires the detection of both spoofing from the start of the test and switching with another person during the test. It is known that existing face authentication technology can perform its task with high accuracy provided that the entire face is captured looking forward with no parts of the face hidden. In actual tests, however, the posture of the test taker tends to change in a variety of ways, so the simple application of face authentication cannot obtain a sufficient level of detection performance.

In response to this problem, NTT DOCOMO developed fraud detection technology combining face authentication technology and tracking technology*1 and an online-testing fraud prevention system applying those technologies. This article reports on this online-testing fraud prevention system and its technologies and on the results of a verification experiment using this system.

2. Online-testing Fraud Prevention System

2.1 Overview

The overall configuration of a system providing online testing is shown in **Figure 1**. Online testing is achieved through an online-testing system, which consists of functions for providing tests such as test-application processing, delivery of test problems, and pass/fail notification, and an online-testing fraud prevention system for preventing fraudulent activity. In this article, we focus on the online-testing fraud prevention system and assume the online-testing system to be an existing system.

The purpose of the online-testing fraud prevention system developed by NTT DOCOMO is to

Table 1 Examples of countermeasures to fraudulent activity in existing online-testing fraud prevention services

	Examples of preventing fraudulent activities
Spoofing	· Identity verification before the test by face authentication [4] [7] · Identity verification by keystroke dynamics [4] · Camera monitoring of test taker by remote proctor [4]-[7]
Cheating	· Camera monitoring of test taker by remote proctor [4]-[6] · Monitoring of test taker's computer screen by remote proctor [5] · Monitoring of test taker's microphone-captured speech by remote proctor [4] [6] · Restricted launching of applications other than those for test taking [4] [6] · Recording of test taker's camera video (to obtain evidence of any fraudulent activity) [4]-[6]

*1 Tracking technology: A technology that tracks a target object in a consecutive sequence of images such as video.

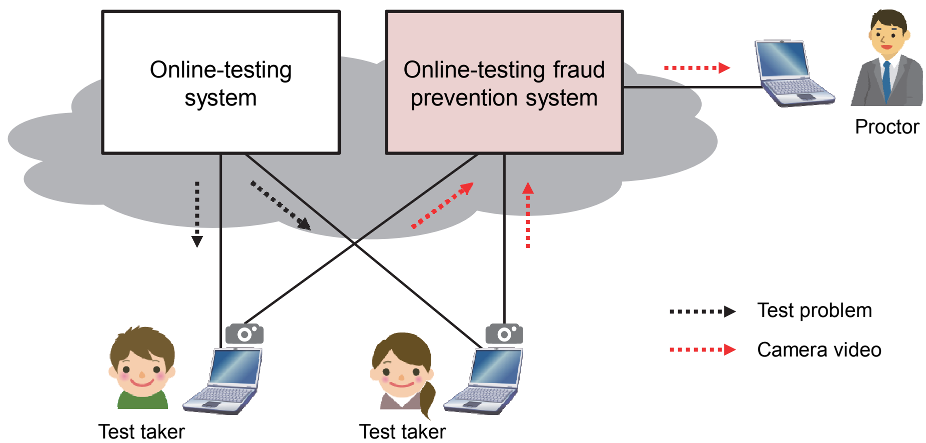


Figure 1 Overall configuration of online-testing system

support test proctors by automatically detecting spoofing through the application of face authentication technology. The following issues arose in the development of this system with regards to enrolling a facial image for cross-checking in face authentication and providing support for test proctors.

(1) Face-image enrollment

- Insuring the identity of the enrolled facial image
- Obtaining a facial image suitable for use in face authentication

(2) Proctor support

- Reliable acquisition of camera video of each test taker during testing
- Automatic detection of spoofing
- Presentation of easy-to-understand detection results to the test proctor
- Saving of evidence of fraudulent activity

The online-testing fraud prevention system consists of a facial-image enrollment function, camera-video acquisition function, video recording function, fraud detection function, and fraud-detection notification function (**Figure 2**). The following describes how each of these functions resolves the above issues.

2.2 System Functions

1) Facial-image Enrollment Function

This function enrolls the test-taker's facial image for cross-checking during face authentication and an image of an identity verification document such as a public identification card for verifying the identity of that facial image. The test taker enrolls a facial image beforehand using a camera connected to the computer being used. At that time, the system guides the test taker in how to capture a clear image of his or her entire face under good

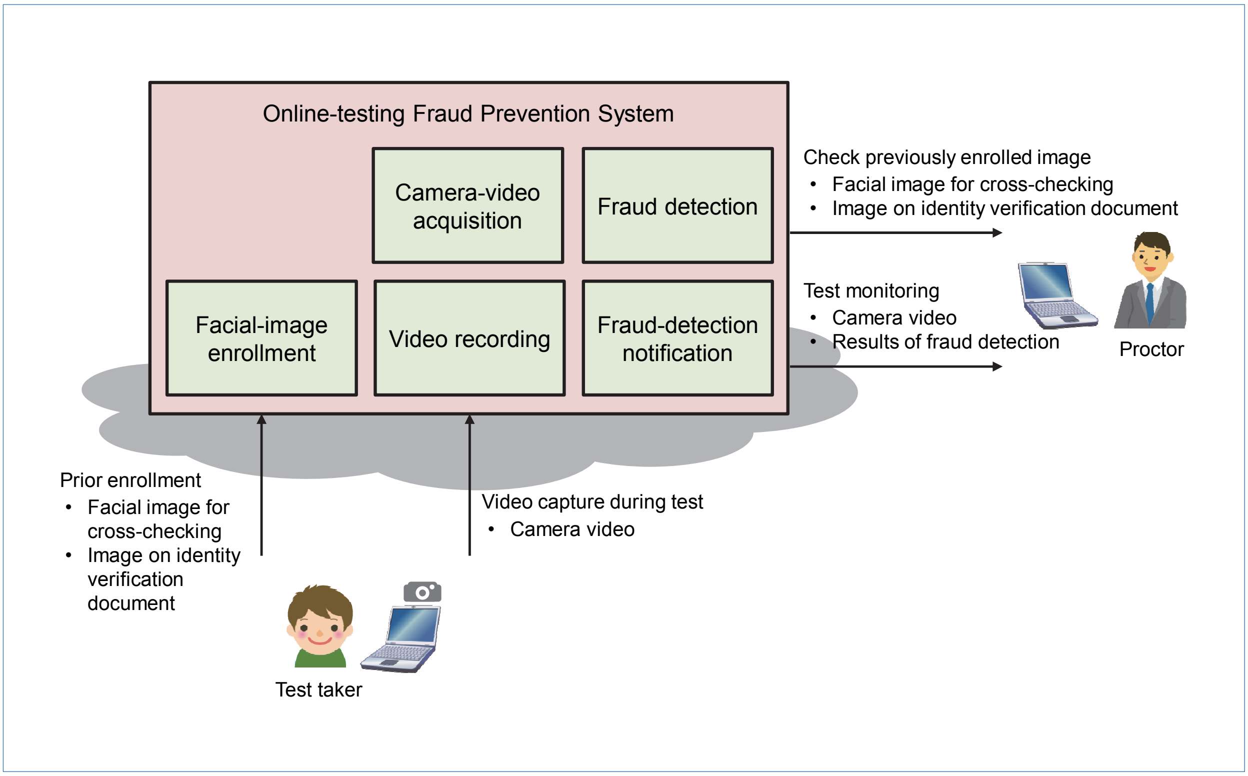


Figure 2 Configuration of online-testing fraud prevention system

lighting conditions. The proctor visually checks whether the enrolled image is suitable enough to identify the test taker in conjunction with the identity verification document and act as a biometric template in face authentication.

2) Camera-video Acquisition Function

This function obtains video of the test taker during the test using a camera connected to the computer. The video is used to automatically detect spoofing during the test and to enable monitoring by the test proctor. If the communications environment should degrade, this video can be obtained through automatic reconnection to the extent possible.

3) Video Recording Function

This function records the video obtained by the camera-video acquisition function. The video is reviewed by the proctor after the test and used as evidence if fraudulent activity has occurred.

4) Fraud Detection Function

This function automatically detects spoofing by applying the fraud detection technology described later to the camera video of the test taker during the test.

5) Fraud-detection Notification Function

This function notifies the proctor of any occurrence of spoofing. Two types of monitoring methods are provided here: monitoring of real-time video simultaneously with the test and viewing of the video after the test. Typical screenshots of the proctor's screen for both methods are shown in **Figure 3** and described below.

- In monitoring by real-time video, a red frame is displayed around the video of that test taker if spoofing is detected and an alert is generated for the proctor (Fig. 3 (a)). This scheme enables a higher number of students that can be simultaneously monitored compared with that by simple visual monitoring.
- In monitoring by recorded video, the occurrence of any spoofing along the band representing the test period is displayed in red at the time corresponding to its detection (Fig. 3 (b)). Clicking on the red portion plays back that video skipping the portion up to that time. This scheme enables the proctor to closely examine the location at which spoofing is suspected thereby shortening checking time significantly compared with viewing the video from the beginning.

3. Overview of Fraud Detection Technology

3.1 Problem of False Positives in Online Testing

The test taker's facial expression and posture can vary in the following ways during testing:

- Change in facial expression (due to frowning, yawning, etc.)
- Change in facial orientation (by looking away, lowering one's eyes, etc.)
- Partial hiding of face (by covering mouth with hand, etc.)
- Partial distortion of face (by resting chin on hand, etc.)

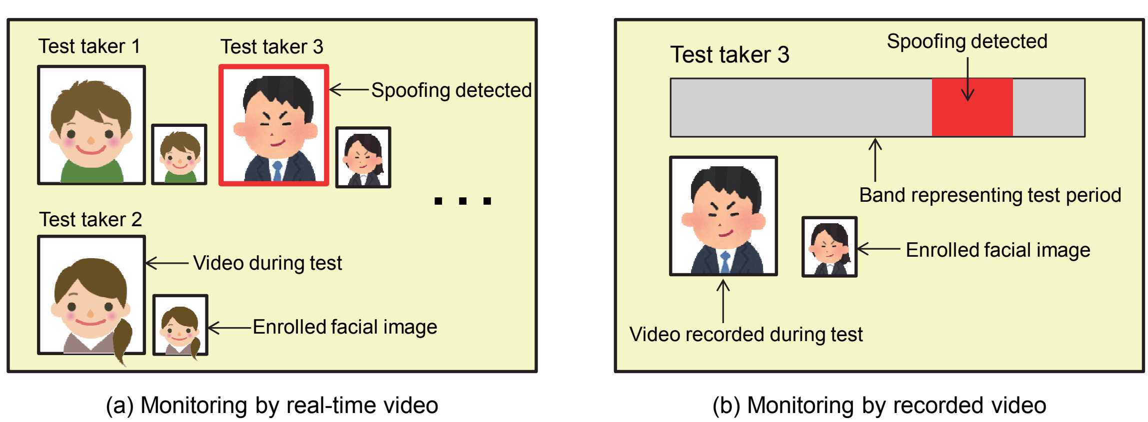


Figure 3 Typical screenshots of proctor's screen

These types of changes can cause the accuracy of face authentication to vary greatly. A problem therefore arises in the simple application of face authentication technology, namely, the frequent occurrence of false positives in which spoofing is judged to have taken place despite the fact that no spoofing actually occurred. Changes in posture and facial expression may be temporary in nature or comparatively long. Given the reality of such changes, finding a way to reduce the occurrence of false positives in fraud detection during online testing has become a key issue. The following describes a method for solving this issue.

3.2 Fraud Detection Technology

The system uses fraud detection technology to perform face authentication against still images extracted at fixed intervals from the video of the test taker as a basis for judging the occurrence of spoofing. However, using the results of face authentication as-is to make a judgment can result in false positives as described above, so the system performs face authentication based on a history of face authentication results and tracking results.

1) Use of Face Authentication History

This fraud detection technology uses a history of face authentication results instead of a single face authentication result at that time to detect spoofing. This has the effect of absorbing temporary fluctuations in face authentication accuracy and decreasing false positives. However, when attempting to absorb long-term changes in posture and facial expression using such a history, a delay in detection can occur or detection of short-term switching with another person becomes impossible. For this reason, a history of face authentication results is used for dealing with short-term changes such as looking away from the computer briefly.

2) Application of Tracking Technology

Fraud detection technology prevents false positives due to long-term changes in posture or facial expression by combining tracking technology with face authentication technology to track the test taker. This tracking technology is used to track the test taker using the last successfully authenticated facial image as the start point. That person is therefore judged to be the target test taker as long as tracking continues to be successful. On the other hand, tracking technology is not authentication technology, so the possibility exists that an impersonator who has taken the place of the test taker will be erroneously tracked and that switching with another person will be missed. Our fraud detection technology prevents the missing of such switching by taking into account the distance moved by the test taker. Additionally, to deal with the problem of a partially concealed face, the similarity of image areas can be calculated by reducing the effect of that concealed portion thereby raising the success rate of tracking.

4. Verification Experiment

Before performing a verification experiment, we evaluated scenarios in which the test taker would intentionally switch with another person. We then prepared an actual qualifying exam in cooperation with a certifying organization and a test delivery operator and performed a total of two verification experiments using our online-testing fraud prevention system in March 2016 and March 2017.

Based on the results of these experiments, we confirmed that the developed fraud detection technology could detect spoofing and curb the occurrence of false positives due to changes in the test taker's posture and facial expressions. We also confirmed that both the real-time-video and recorded-video monitoring methods could provide the results of spoofing detection in an easy-to-understand manner thereby providing effective support for test proctors.

5. Conclusion

Focusing on the automatic detection of spoofing in online testing, this paper presented the technical issues involved, described an online-testing fraud prevention system and associated fraud detection technologies for solving those issues, and demonstrated the effectiveness of the system through a verification experiment. In future research, we plan to study the automatic detection of cheating in online tests.

REFERENCES

[1] Coursera Website.
https://www.coursera.org/

[2] edX Website.
https://www.edx.org/

[3] JMOOC Website.
https://www.jmooc.jp/

[4] KRYTERION Website.
https://www.kryteriononline.com/

[5] ProctorU Website.
https://www.proctoru.com/

[6] RPNow Website.
https://www.psionline.com/platforms/rpnow/

[7] H. Kawahara: "Student Authentication for Course Credit in Distance Learning," Journal of Multimedia Education Research, Vol.7, No.1, 2010.
http://www.code.ouj.ac.jp/media/pdf/vol7no1_shotai7_071.pdf

Technology Reports

Drone | Spherical Display | PoV Display

Spherical Drone Display

Research Laboratories Wataru Yamada Hiroyuki Manabe Daizo Ikeda
Innovation Management Department Kazuhiro Yamada

The spherical drone display is the world's first airborne spherical display that can display video images in any location while flying. Until now, equipping a drone with a high resolution, large screen display has been problematic due to weight and aerodynamic considerations. This device uses persistence of vision effect display technology and achieves a spherical display that flies stably.

1. Introduction

Magical technologies that create airborne images have long been featured in the world of science fiction, and there have been many studies on how to create these technologies in the real world. These include technologies known as Augmented Reality (AR)*1 to display images created by combining real scenes and computer graphics [1], image projection technologies using floating magnetic spheres [2], and laser technologies that generate plasma in the air [3].

Another of these technologies gaining a lot of attention in recent years involves drones fitted with light sources such as the Light Emitting Diodes (LEDs) to display the airborne images. For example, Intel's "Shooting Star" system uses and simultaneously controls several hundred microdrones fitted with a high luminosity LEDs to produce a giant image floating in the air [4]. Another similar system is the Japanese company MicroAd's SKY MAGIC image display system using drone groups [5]. These technologies are pioneering a new area of image display by establishing new methods to float images in the air. However, these current systems can only have one or several LED lights fitted to each drone, and when flying in groups, the number that can be simultaneously controlled is limited

©2018 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

*1 AR: Technology for superposing digital information on real-world video in such a way that it appears to the user to be an actual part of that scene.

from several tens to hundreds of devices, meaning displays are limited to low resolution images such as simple logos and shapes.

Making use of the flying abilities of drones, NTT DOCOMO has developed a "spherical drone display" by developing improvements to the expressive capabilities of individual drones to make it easier to more freely display images. The spherical drone display enables, for example, dynamic performances that move around in the air at concerts or live venues, and new advertising display medium that flies around venues.

This article describes an overview and the structure of the spherical drone display, issues with implementation, the advantages of this technology, and future improvements.

2. Spherical Drone Display

2.1 Overview

It has been technically difficult to mount high resolution, wide image display screens on drones without adversely affecting their flying abilities. This is because mounting dense arrays of many LEDs or large displays to enable high expressivity can inhibit drone airflow and increase body weight, making it difficult to fly. In other words, with drones, there is a trade-off between flying ability and visual expression capability.

To counter this, NTT DOCOMO has applied a "spherical display technology" to drones to enable high display capabilities without losing flight capabilities. Spherical displays make use of a phenomenon called the persistence of vision (PoV) effect, in which the light and image seem to remain after they have been viewed. Hence, by flashing a light source and moving it, the PoV effect enables images to appear to the viewer along the trajectory of the moving light source. NTT DOCOMO has combined a drone with a spherical display using this PoV display technology to create an omnidirectional spherical drone display, as shown in **Figure 1** (a) and (b) [6].

Figure 1 (a)　A flying spherical drone display

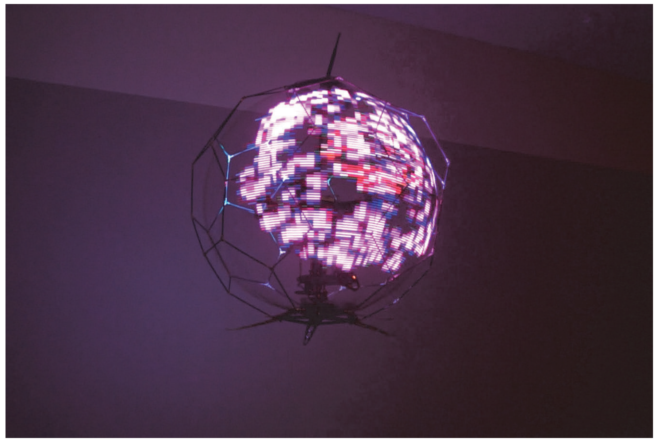


Figure 1 (b)　An image of fireworks displayed on the device

2.2 Device Structure

Figure 2 and **3** show the appearance and structure of the spherical drone display. The device is 88 cm in diameter, and weighs approximately 4.5 kg. The spherical drone display consists of the following three main parts, listed in order from the inside outward.

(1) The drone that produces thrust to enable flight
(2) PoV display to produce the spherical image
(3) A protector for the propellers and LEDs

- The drone is positioned in the middle of the spherical drone display. To fly, the device has four motors attached to 13-inch propellers, and two auxiliary motors and propellers. The four motor-propeller sets are used by the drone for flight, and the two auxiliary motor-propeller sets are used to counter the reaction force of the rotating PoV display, described later. For rigidity and light weight, the frame is mainly made of machined carbon.
- The spherical display consists of eight strips of arc-shaped LED tape*2 encompassing the drone and a mechanism to rotate them. There are 144 LEDs on each tape. The eight LED tapes rotate three times per second, and the LEDs are synchronized to display the spherical image. The LEDs flash 136 times per rotation, which produces an overall resolution of vertical (half-period) 144 x horizontal (one period) 136 pixels. Also, there is a frame rate of 24 fps (frames per second) for the three rotations per second of the eight LED strips.
- The protector is on the outside of the PoV display, to prevent the drone propellers and rotating LEDs from colliding with people or obstacles. The protector must be lightweight and strong, and is designed as a multi-surface shape called a “truncated icosahedron” made of carbon pipes and aluminum joints. The protector

*2 LED tape: An array of LEDs positioned on a tape-like substrate. This technology uses 144 LEDs arranged in a single line on a 1 m long tape.

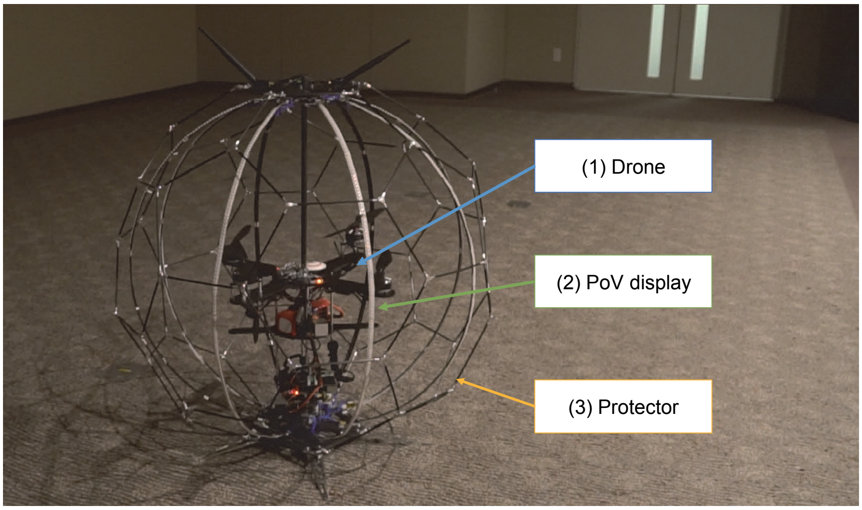


Figure 2　Spherical drone display appearance

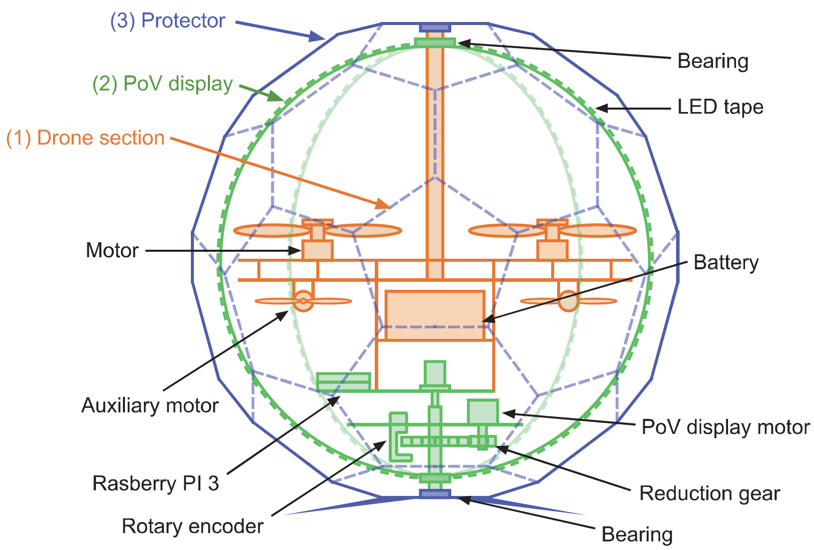


Figure 3　Schematic diagram of spherical drone display structure

does not rotate during display, because the PoV display section rotates independently of the protector and drone sections.

2.3 Challenges and Solutions

Even though the idea behind this technology is simple, there are a range of technical difficulties in its achievement. These difficulties include various challenges such as balancing weight with strength, and high-speed LED control. Particularly, due to the rotation of the LEDs to produce the image, there is a reaction force from the rotation on the drone, which would cause the drone to rotate and make it uncontrollable. To solve this issue, we added two auxiliary motor-propeller sets which rotate to counter the rotational speed of the LEDs to eliminate the reaction force of the PoV display.

2.4 Advantages

The spherical drone display is a combination of the PoV display and a drone, and has the following advantages.

(1) Although the device has a large display surface encompassing the drone and a high apparent resolution, weight and interference with drone airflow are minimized because the actual display surface is only a few narrow arc-shaped LED strips. This has enabled mounting high-resolution, large-screen display on a drone, which was previously difficult.

(2) Because the display is a sphere, it can be viewed from any direction. This is a particularly important advantage because it enables images such as stage presentations or advertising to be delivered in all directions in a venue.

(3) Visibility is high because the drone section does not block the image, and safety is ensured, because all parts including the drone drive mechanism such as motors, propellers and rotating LEDs etc. are housed inside the protector.

(4) High luminosity LED light sources enable an extremely bright display. However, because images are difficult to recognize in direct sunlight etc., users should consider the place and surrounding light conditions when using the device.

(5) PoV displays can display images across the entire surface as shown in Fig. 1 (a), or display partial images as shown in Fig. 1 (b). In addition, because parts that don't display anything appear invisible in dark places, it's possible to use the device to suddenly display images in the air as if they appear from nowhere.

2.5 Future Improvements

There is an issue, because drone mechanisms create a lot of noise when flying. However, it's possible to mask the noise when using the devices at music events such as drone shows, live performance or concerts if the music is sufficiently loud.

The current flying time is dependent on the size of the battery, which is approximately five minutes. Nevertheless, as lighter and lighter batteries are currently under development, flying time is expected to reach approximately 10 to 15 minutes, which will be sufficient to fly for one or two songs during a show or performance.

While the resolution of this display is high compared to drone displays using conventional LEDs, the display is very rough compared to a typical display. This issue could be improved by using specialized boards with higher densities of LEDs, or by positioning adjacent LED tapes so their LEDs are between each other.

3. Conclusion

This article has described a spherical drone display that floats in the air and is capable of omnidirectional image display. This technology featured on several hundred television stations, newspapers and Web news sources etc. both in Japan and around the world since the release of information through the DOCOMO official Web site on April 17, 2017. Then, the device was flown in public for the first time at "NTT ULTRA FUTURE MUSEUM 2017" at the Niconico Chokaigi 2017 held on April 29 - 30. This two-day event attracted several thousand visitors, and the demonstration of the device attracted a great deal of attention. Please follow the link at reference [7] to see a video of this technology. As well as making improvements such as flight time and resolution, we will work towards commercializing the technology for new aerial performances and advertising solutions by incorporating autonomous and group flight technologies.

REFERENCES

[1] T. P. Caudell and D. W. Mizell: "Augmented reality: an application of heads-up display technology to manual manufacturing processes," In Proc. of HICSS '92, Vol.2, pp.659-669, 1992.

[2] J. Lee, R. Post and H. Ishii; "ZeroN: mid-air tangible interaction enabled by computer controlled magnetic levitation," In Proc. of UIST '11. pp.327-336. 2011.

[3] H. Kimura, T. Uchiyama and H. Yoshikawa: "Laser produced 3D display in the air," In Proc. of ACM SIGGRAPH 2006 Emerging technologies, No.20, 2006.

[4] Intel: "Intel-based Drone Technology Pushes Boundaries. (2016)," Retrieved Mar. 2017.
http://www.intel.com/content/www/us/en/technologyinnovation/aerial-technology-overview.html

[5] SKY MAGIC Web site.
https://magic.microad.co.jp/skymagic/

[6] W. Yamada, K. Yamada, H. Manabe and D. Ikeda: "iSphere: Self-Luminus Spherical Drone Display," In Proc. of UIST '17, 2017.

[7] DOCOMO R&D Magazine: "Images in the Air with Drones, Technologies to Stimulate the Five Senses, the Future of Entertainment at Niconico Chokaigi 2017," (In Japanese).
http://style20.jp/rd/article/article_32/

Topics

Efficient Corporate User Smartphone Kitting with Android Kitting Tool

Solution Service Department Takeshi Ishii Shinya Hasegawa

When delivering Android™*1 smartphones to corporate customers, initial settings unique to each company called "kitting" must be performed for each handset. For example, these settings include screen lock settings, Wi-Fi®*2 connection settings, adding and settings of applications and settings to position applications on the home screen. Kitting is performed in advance so that customers can use the Android smartphones in their business as soon as they are delivered, and it is thus a necessary process in corporate business which improves customer satisfaction levels.

The number of kitting work items is on the rise due to the many settings for Android smartphones and the variety of customer usage methods and control items of Mobile Device Management (MDM)*3 are expanding. Depending on the case, it can take several hours to kit one device, or several months to kit all devices to be delivered to a company. The timing of kitting work for each project is unclear, and is not possible to increase the number of staff to cope without planning. Additionally, many delivery cycles are short, which makes shortening work time an urgent matter. Also, accuracy must be improved to prevent settings errors and raise productivity.

To solve these issues, NTT DOCOMO has developed and deployed an Android kitting tool (hereinafter referred to as "AKT") for its in-house PCs to automate kitting work. AKT has been reducing the amount of work, shortening delivery time and improving accuracy.

This article describes the AKT development.

1) The Initial AKT

At first, we designed AKT as a tool to automatically set approximately 50 items, using an Application Programming Interface (API)*4 method and a UI Automator*5 method (**Figure 1**).

(1) API method

In Android smartphones, there are APIs available for applications and their settings, and terminal settings. For example, screen brightness settings and so forth can be changed from other applications etc. via connection to the screen settings API (**Figure 2**). Not necessarily all settings are accessible through the API, but all settings available with the API are performed.

(2) Mechanism with the UI Automator method

UI Automator can acquire on-screen information, and can simulate key events such as tap and scroll using commands. For example,

©2018 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

*1 Android™: A trademark or registered trademark of Google, Inc.

*2 Wi-Fi®: A registered trademark of the Wi-Fi Alliance.

*3 MDM: A method of managing and controlling the functions and applications of smartphones supplied to staff of a company etc.

key event commands can be used to start downloading a particular application, analyze the characters displayed when download finishes, and then execute the subsequent action. UI Automator is used for kitting items not executable through the API (**Figure 3**).

(3) Issues with the initial AKT

Because there was only a limited number of items that could be set with the initial AKT, not all items required for kitting

*4 API: Interfaces specified for communications between different software.

*5 UI Automator: A test tool program provided by Google, Inc.

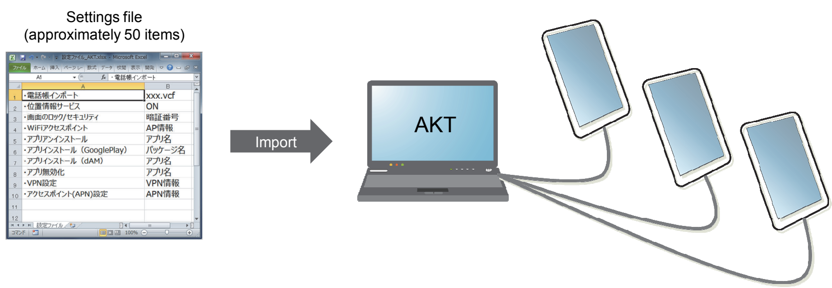


Figure 1　The Initial AKT

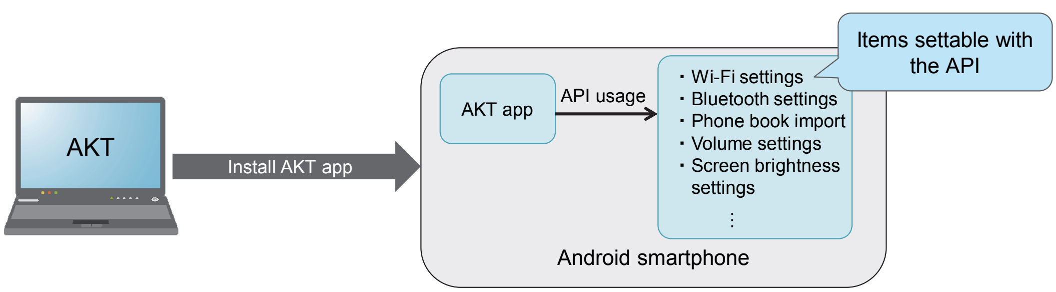


Figure 2　API method

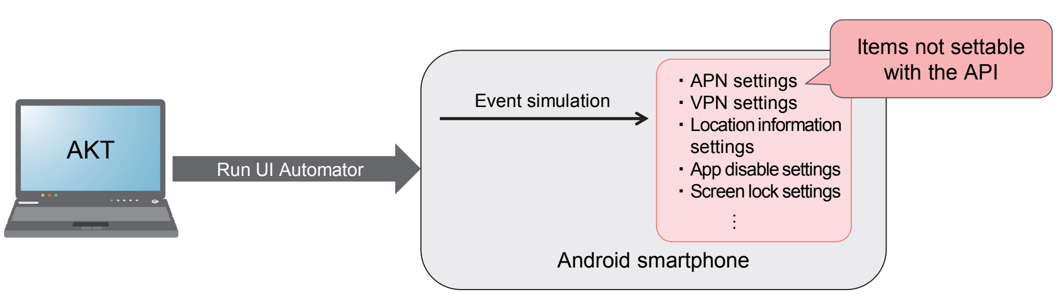


Figure 3　UI Automator method

were covered. Furthermore, because the scope varied depending on the details of kitting, usage of this system required judgments to identify levels of efficiency and effectiveness. Hence, the system was not always used and did not become established in-house.

Also, there was the operational issue of additional development required with AKT so that UI Automator could handle new models, which entailed long development periods and difficulty in reducing development costs.

2) AKT Improvements

We broadened the scope of AKT functions so that it can be used more efficiently with a wider range of projects.

(1) Operation scenario function development

Because the initial AKT only offered limited number of settings items, we developed and added a function for operation scenarios to pseudo-replay Android smartphone operations.

With this function, an Android smartphone is first connected to the AKT and kitted, which is recorded in AKT as a scenario. Next, multiple unkitted Android smartphones are connected to AKT, and the scenario is run to automatically reproduce the same operations as the Android smartphone that was connected first (**Figure 4**).

Firstly, with this function, we assumed use of getevent (acquire touch event)/sendevent (send touch event) operations available with Android Debug Bridge (ADB)*6, but as a result of testing, the replay processing speed was slow, and it was clear that the actions were not the same as they were recorded. This was caused by sendevent processing delays due to internal processing congestion with sendevent processing for each touch event (**Figure 5** (a)). Therefore, to prevent internal processing congestion, we changed to write the touch event acquired with getevent directly in the Android system using Android Native Development Kit (Android NDK)*7, without using sendevent. With this change, we were able to eliminate processing latency with replay and confirm correct operation of the function (Fig. 5 (b)).

The inclusion of the operation scenario function makes it possible to automate all

*6 ADB: A development tool program provided by Google, Inc., that enables Android smartphone control from PC.

*7 AndroidNDK: A development kit provided by Google, Inc. for Android application development in the C/C++ languages. This software has a high processing speed because it is not run via an execution environment for Java® language applications. Oracle and Java are registered trademarks of Oracle Corporation and its subsidiaries and related companies in the United States and other countries.

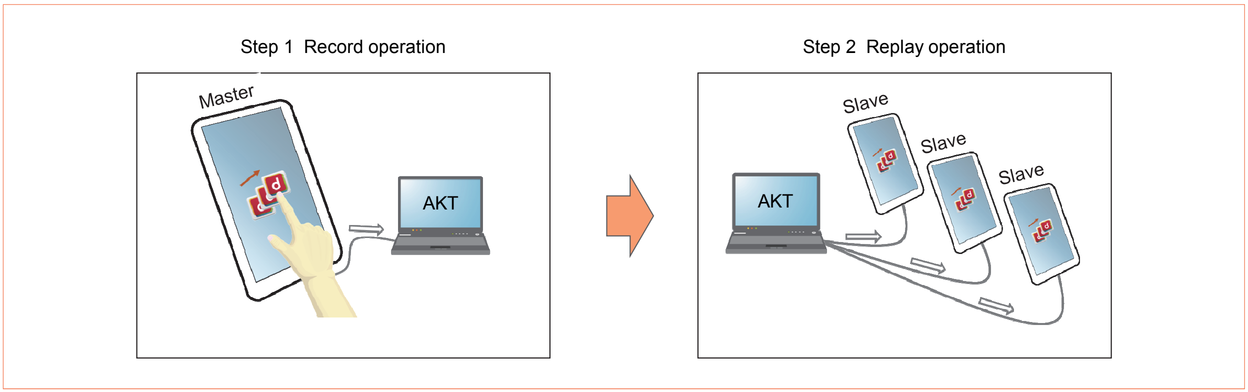


Figure 4 Image of operation scenario function

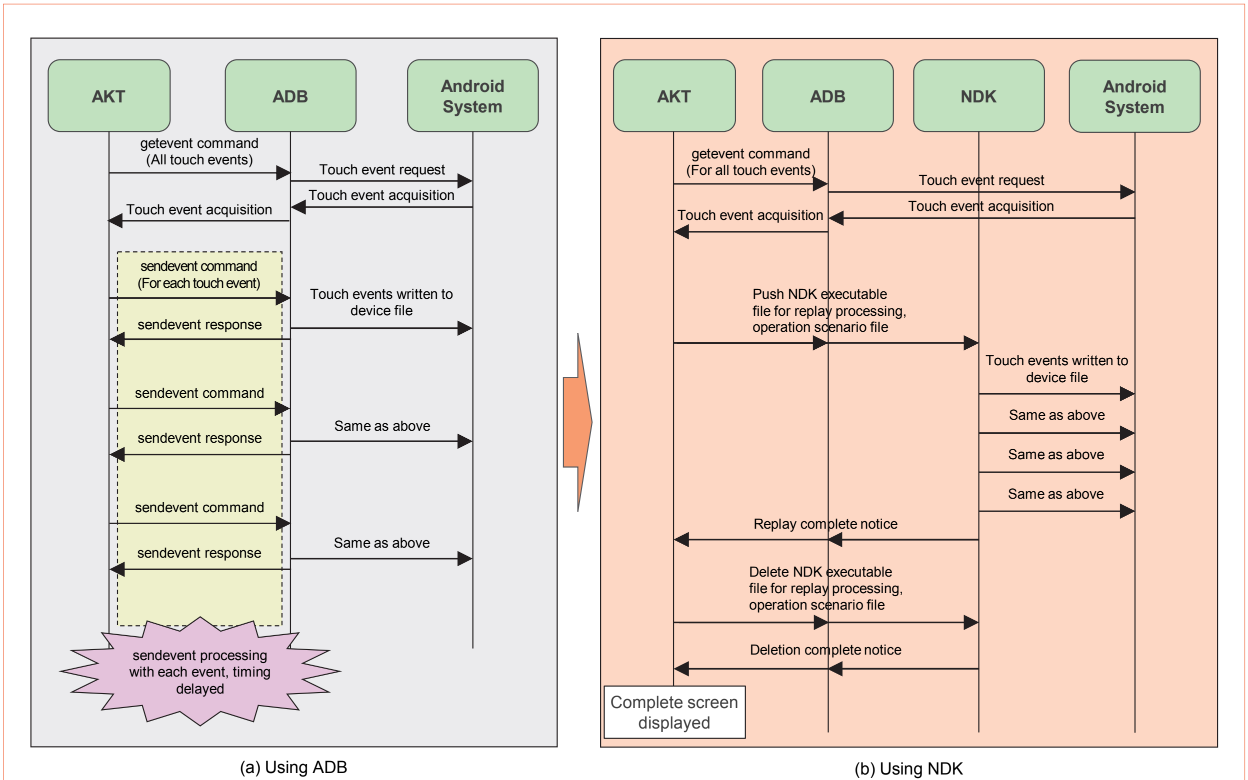


Figure 5　Operation scenario function sequence

operations, and greatly expands the AKT usage scope. Also, this simple-to-use function makes the tool easy to use for people who are troubled by difficult operations.

(2) Text paste function development

We also developed a function to automatically input different values into multiple Android smartphones.

This function enables input of values specified for individual handsets with AKT by loading a list of values corresponding to Android smartphone serial numbers in advance.

This function, in combination with the aforementioned operation scenario function, enables automation of inputs such as individual IDs which formerly required settings for each unit, and makes it possible to automate almost all operations required for kitting (**Figure 6**).

(3) API method expansion

We expanded items settable with APIs in cooperation with terminal manufacturers. Unlike UI Automator, in many cases, APIs are not model-dependent, and can be used with any model. By replacing the UI Automator with APIs, we were able to minimize the amount of AKT development required to support new models, which sped up development and reduced development costs.

AKT became established in-house because we broaden the scope of kitting that can be automated. This has greatly contributed to reducing work time, shortening delivery times and improving accuracy with kitting work.

Improving functionality greatly raised effectiveness. The following two points were key to this success.

- We held close and patient discussions with

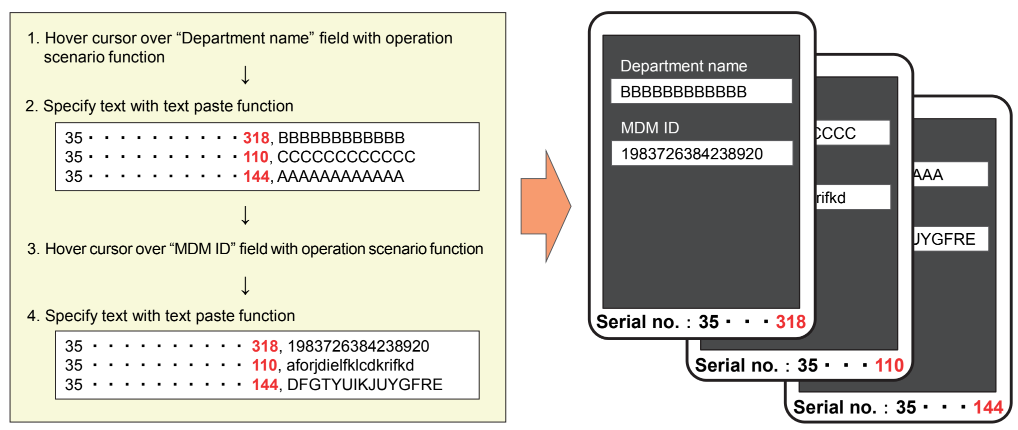


Figure 6　Text paste function

those in the workplace to accurately understand needs and determine the required functionality.

- To develop the exact functionality required, we avoided stereotypes and conceived new ideas from the ideal.

Thoroughly focusing on these two aspects enabled us to produce the current AKT and achieve more efficient kitting work. We will continue to advance AKT to further reduce workloads in the workplace.

NTT DOCOMO
Technical Journal　Vol.19　No.4

Editorship and Publication
NTT DOCOMO Technical Journal is a quarterly journal edited by NTT DOCOMO, INC. and published by The Telecommunications Association.

Editorial Correspondence
NTT DOCOMO Technical Journal Editorial Office
R&D Strategy Department
NTT DOCOMO, INC.
Sanno Park Tower
2-11-1, Nagata-cho, Chiyoda-ku, Tokyo 100-6150, Japan
e-mail: dtj@nttdocomo.com

Copyright
ⓒ 2018 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.