

# IWIN2020



## **International Workshop on Informatics**

Proceedings of  
International Workshop on Informatics

September 10-11, 2020  
Virtually Wakayama



Sponsored by Informatics Society





# IWIN2020



## **International Workshop on Informatics**

Proceedings of  
International Workshop on Informatics

September 10-11, 2020  
Virtually Wakayama



Sponsored by Informatics Society

Publication office:

Informatics Laboratory

3-41, Tsujimachi, Kitaku, Nagoya 462-0032, Japan

Publisher:

Tadanori Mizuno, President of Informatics Society

ISBN:

978-4-902523-47-8

Printed in Japan

# Table of Contents

## **Session 1: Multimedia**.....1

**( Chair: Katsuhiko Kaji ) ( 9:40 - 10:35, Sep. 10 )**

- (1) A Study on Presentation of Viewers' Interests based on POV Analysis in Mobile 360-degree Internet Live Broadcasting.....3  
Masaya Takada, Yoshia Saito
- (2) A Privacy-oriented Video Distribution Platform for Public Camera Systems.....9  
Satoru Matsumoto, Tomoki Yoshihisa, Tomoya Kawakami, Yuuichi Teranishi

## **Keynote Speech 1**.....16

**( 10:45 - 11:25, Sep. 10 )**

- (I) Digital Transformation with Trust .....17  
Dr. Rieko Yamamoto, Fellow at the Fujitsu Laboratories Ltd.

## **Session 2: Sensing and Analysis.....41**

**( Chair: Yoshitaka Nakamura ) ( 13:00 - 15:10 , Sep. 10 )**

- (3) To Estimate a Specific Position Related to an Event .....43  
Takuma Toyoshima, Takuo Kikuchi, Masaki Endo, Shigeyoshi Ohno,  
Hiroshi Ishikawa
- (4) Anomaly detection in FA equipment using an interaction model.....49  
Hiroaki Ando, Yusuke Iwatsuki, Daiki Hibi, Kazuhiko Tsutsui, Satoshi Aoki,  
Katsuhiro Naito, Naoya Chujo, Tadanori Mizuno, and Katsuhiko Kaji
- (5) Object Extraction Method from Mobile Camera Videos Using Optical Flow .....55  
Tsukasa Kudo
- (6) Automatic detection of tourist spots and best-time estimation using social network services  
..... 65  
Munenori Takahashi, Masaki Endo, Shigeyoshi Ohno, Masaharu Hirota,  
Hiroshi Ishikawa
- (7) Radiation Energy of Classical and Nonclassical Vocal Styles .....73  
Orie Abe, Rolf Bader

## **Session 3: Systems and Services.....81**

**( Chair: Yoshia Saito ) ( 15:20 - 17:05, Sep. 10 )**

- (8) Process Improvement Method of Progress Meeting ..... 83  
Akihiro Hayashi
- (9) Ad Hoc Transmission Algorithm for Low-Price Rice Field Server .....89  
Mikiko Sode Tanaka, Yuki Okumura, Sota Tatsumi, Shogo Ishii, Tatsuya Kochi
- (10) Java Model Checking: Improvement of the Understanding of Counterexample.....95  
Chellet Marwan Bernard Hassan, Shinpei Ogata, Kozo Okano
- (11) Enhancing RTK-GNSS Infrastructure in Snowy and Mountain Region Through Rule-  
Based Base Station Assignment Approach.....101  
Bhagawan Rokaha, Bishnu Prasad Gautam, Tomoya Kitani

## **Session 4: Network and Security.....109**

**( Chair: Tomoki Yoshihisa ) ( 9:20 - 11:30, Sep. 11 )**

- (12) A Proposal of Autonomous Control of Server Relocation for Fog Computing Systems· 111  
Kouki Kamada, Hiroshi Inamura, Yoshitaka Nakamura
- (13) On Improving Efficiency of CSMA/CA with RSSI-based Control-frame Detection .... 117  
Yoshito Umezawayand, Takuya Yoshihiro
- (14) A Study on System Architecture of Smart Lock Based on Authentication of Door  
Knocking Motion Using Machine Learning..... 123  
Kakeru Nakabachi, Yoshitaka Nakamura, Hiroshi Inamura
- (15) A Proposal of Personal Authentication Method Based on Eye Movement Trajectory with  
Fixation and Saccade Features ..... 133  
Takumi Fujimoto, Yoh Shiraishi
- (16) Hybrid CSMA/CA for Sub-1 GHz Frequency Band Coexistence of IEEE 802.11ah and  
IEEE 802.15.4g ..... 139  
Yukimasa Nagai, Jianlin Guo, Takenori Sumi, Philip Orlik, Hiroshi Mineno

## **Session 5: ITS**.....145

**( Chair: Tomoya Kitani ) ( 13:00 - 14:45, Sep. 11 )**

- (17) A Route Guidance Method for Vehicles to Reduce Traffic Congestion Considering Delay  
in Destination Arrival Time ..... 147  
Yusuke Matsui, Takuya Yoshihiro
- (18) Examination of Environment Recognition Method for Autonomous Driving on  
Undeveloped Road ..... 153  
Kazumasa Kamitani, Naoya Chujo
- (19) Memory-Saving Software Update Method for In-Vehicle ECU ..... 161  
Kazuki Someya, Yoshiaki Terashima, Shunsuke Sugimoto, Ryoza Kiyohara
- (20) A Study of Patrol System by Automated Robotic Car..... 169  
Yuto Nagai, Yuya Sawano, Ryoza Kiyohara

## **Session 6: Industrial and Agricarutural**

## **Applications**.....175

**( Chair: Hiroshi Inamura ) ( 14:55 – 16:40, Sep. 11 )**

- (21) High accuracy synchronized spindle/Servo control in the CNC Equipment ..... 177  
Kazuhiko Tsutsui, Katsuhiko Kaji, Katsuhiro Naito, Naoya Chujo, Tadanori Mizuno
- (22) Development of a Yield Prediction Model Generation Process for Fruits and Vegetables in  
a Plant Factory ..... 183  
Yuki Todate, Michiko Oba, Mitsuru Takamori
- (23) Proposal for Method of Efficient Building Facility Management linking Human Sense and  
Information of BEMS ..... 191  
Fumiaki Kimura, Fuyuki Sato, Yoshihito Endo, Takeyuki Kimura, Shinji Kitagam
- (24) Regression Combined with Random Forest for Wagyu Quality Prediction ..... 197  
Shingo Tsukamoto, Haruka Ikegami, Tamako Matsuhashi, Kazuya Matsumoto,  
Takuya Yoshihiro

## Message from the General Chairs



It is our great pleasure to welcome all of you to Wakayama, Japan (virtually because of the COVID-19 pandemic), for the 14th International Workshop on Informatics (IWIN 2020). This workshop has been held annually by the Informatics Society. Since 2007, the workshops were held in Naples in Italy, Wien in Austria, Hawaii in the USA, Edinburgh in Scotland, Venice in Italy, Chamonix in France, Stockholm in Sweden, Prague in Czech Republic, Amsterdam in Netherlands, Riga in Latvia, Zagreb in Croatia, Salzburg in Austria, and Hamburg in Germany respectively.

In IWIN 2020, 24 papers were accepted after peer reviewing by the program committee. Based on the papers, six technical sessions were organized in a single-track format, which highlighted the latest research results in the areas such as Multimedia, Sensing and Analysis, Systems and Services, Network and Security, Intelligent Transport System (ITS), Industrial and Agricultural Applications. IWIN2020 will also welcome one keynote speaker: Dr. Rieko Yamamoto, Fellow of Fujitsu Laboratories Ltd. We really appreciate her participation in the workshop.

We would like to thank all the participants and contributors who made the workshop possible. It is indeed an honor to work with a large group of professionals around the world for making the workshop a great success. We are looking forward to seeing you all in the workshop. We hope you enjoy IWIN 2020.

September 2020

Hiroshi Ishikawa  
Takuya Yoshihiro

# Organizing Committee

## General Co-Chairs

Hiroshi Ishikawa (Tokyo Metropolitan University, Japan)

Takuya Yoshihiro (Wakayama University, Japan)

## Steering Committee

Hitoshi Aida (Tokyo University, Japan)

Toru Hasegawa (Osaka University, Japan)

Teruo Higashino (Osaka University, Japan)

Tadanori Mizuno (Aichi Institute of Technology, Japan)

Jun Munemori (Wakayama University, Japan)

Yuko Murayama (Tsuda College, Japan)

Ken-ichi Okada (Keio University, Japan)

Norio Shiratori (Chuo University / Tohoku University, Japan)

Osamu Takahashi (Future University Hakodate, Japan)

## Program Chair

Katsuhiro Kaji (Aichi Institute of Technology, Japan)

## Financial Chair

Tomoya Kitani (Shizuoka University, Japan)

## Publicity Chair

Yoshitaka Nakamura (Future University Hakodate, Japan)

## Program Committee

Akihito Hiromori (Osaka University, Japan)

Akira Uchiyama (Osaka University, Japan)

Chiaki doi (NTT DOCOMO, Inc., Japan)

Fumiaki Sato (Toho University, Japan)

Gotoh Yusuke (Okayama University, Japan)

Hideyuki Takahashi (Tohoku Gakuin University, Japan)

Hiroaki Morino (Shibaura Institute of Technology, Japan)

Hironobu Abe (Tokyo Denki University, Japan)

Hiroshi Mineno (Shizuoka University, Japan)

Hiroshi Sugimura (Kanagawa Institute of Technology, Japan)

Hiroshi Yoshiura

(The University of Electro-Communications, Japan)

Hisao Fukuoka (Tokyo Denki University, Japan)

Katsuhiro Naito (Aichi Institute of Technology, Japan)

Kei Hiroi (Kyoto University, Japan)

Keiichi Abe (Kanagawa Institute of Technology, Japan)

Ken Ohta (NTT DOCOMO, Inc., Japan)

Kozo Okano (Shinshu University, Japan)

Makoto Imamura (Tokai University, Japan)

Masaaki Shirase (Future University Hakodate, Japan)

Masaji Katagiri (NTT DOCOMO, Inc., Japan)

Masakatsu Nishigaki (Shizuoka University, Japan)

Masaki Endo (Polytechnic University, Japan)

Masashi Saito (Kanazawa Institute of Technology, Japan)

Minoru Kobayashi (Meiji University, Japan)

Naoya Chujo (Aichi Institute of Technology, Japan)

Ryozo Kiyohara (Kanagawa Institute of Technology, Japan)

Shigemi Ishida (Kyusyu University, Japan)



Shinichiro Mori (Chiba Institute of Technology, Japan)  
Shinji Kitagami (Fukui University of Technology, Japan)  
Takaaki Umedu (Shiga University, Japan)  
Takaya Yuizono  
(Japan Advanced Institute of Science and Technology)  
Takayasu Yamaguchi (Akita Prefectural University, Japan)  
Tetsushi Ohki (Shizuoka University, Japan)  
Tetsuya Shigeyasu  
(Prefectural University of Hiroshima, Japan)  
Tomoki Yoshihisa (Osaka University, Japan)  
Tomoo Inoue (University of Tsukuba, Japan)  
Tomoya Kitani (Shizuoka University, Japan)  
Tomoyuki Yashiro (Chiba Institute of Technology, Japan)  
Tsukasa Kudo  
(Shizuoka Institute of Science and Technology, Japan)  
Yoshia Saito (Iwate Prefectural University, Japan)  
Yoshiaki Terashima (Soka University, Japan)  
Yoshinobu Kawabe (Aichi Institute of Technology, Japan)  
Yoshitaka Nakamura (Future University Hakodate, Japan)  
Yu Enokibori (Nagoya University, Japan)  
Yusuke Ichikawa  
(NTT Service Evolution Laboratories, Japan)  
Yusuke Takatori (Tokyo University of Science, Japan)



Session 1:  
Multimedia  
( Chair: Katsuhiko Kaji )



# A Study on Presentation of Viewers' Interests based on POV Analysis in Mobile 360-degree Internet Live Broadcasting

Masaya Takada\* and Yoshia Saito\*

\*Graduate School of Software and Information Science, Iwate Prefectural University, Japan  
g236q002@s.iwate-pu.ac.jp, y-saito@iwate-pu.ac.jp

**Abstract** - 360-degree Internet live broadcasting is a live broadcast using an omnidirectional camera. With the advent of various inexpensive omnidirectional cameras, this service is now available to general users. In addition, with the development of the Internet infrastructure, users are able to conduct the 360-degree Internet live broadcasting outdoors. This service can now be used as a mobile service. The features of this service have the ability to provide a greater amount of information than those of conventional broadcasting and a greater degree of freedom in viewing direction (POV: Point Of View). On the other hand, the 360-degree Internet live broadcasting services do not have the ability for the broadcaster to know the viewers' POV. The role of gaze information in remote communication is very important, as it shows the focus of the conversation and the object of interest. In other words, if the broadcaster cannot be aware of the viewers' POV, it is not possible to respond appropriately to the viewer's comments. For this problem, we have analyzed the POV and created an algorithm to detect viewers' interests. The algorithm used characteristics about the viewer's viewing behavior. It could detect significant POV changes which represent viewers' interests with 89.76% accuracy. In this paper, we show an experimental result to evaluate the effect of presenting the algorithm outputs to the broadcaster.

**Keywords:** 360-degree Internet live broadcasting, Viewers' POV.

## 1 INTRODUCTION

Interest in Internet live broadcasting services is increasing year by year, and many users enjoy real-time communication through live broadcasting services. Since Internet live broadcasting allows the broadcaster and viewers to enjoy real-time communication, it is also used as a communication tool. YouTube and some Internet live broadcasting services, such as Facebook, support for omnidirectional cameras. This service is called 360-degree Internet live broadcasting, and it is a groundbreaking service that provides more information about around of the broadcaster than that of the conventional broadcasting. For example, the broadcaster who goes sightseeing can provide the entire space of tourist spot using the 360-degree internet live broadcasting. However, unlike conventional broadcasting, the 360-degree Internet live broadcasting uses an omnidirectional camera. It is difficult for the broadcaster to intuitively grasp the viewer's viewing range (POV: Point Of View). It is necessary to present the viewers' POV to the broadcaster for smooth communication. This is because

previous researches about remote communication have shown that the communicator's gaze indicated the target of interest or center of the topic [1][2]. The lack of information may lead to discrepancies in communication. The current services only support comments, when the viewers communicate with the broadcaster. The broadcaster has to respond based on the text information. Therefore, we address this issue by using the viewers' POV to facilitate communication by adding the ability to analyze and present it to the broadcaster.

The contributions of this paper are summarized as follows:

- We developed an algorithm to detect viewers' interests based on the characteristics of viewers' POV.
- We clarified the effects of the proposed algorithm when the analyzed results of the viewers' POV are presented to the broadcaster.

The rest of this paper is organized as follows. Section 2 describes the use cases and the advantage of our proposal. Section 3 describes the related work and discuss the necessity of this study. Section 4 describes the hypotheses about the viewer's viewing behavior and their testing. Section 5 describes an overview of the algorithm and its preliminary experiment. Chapter 6 describes improvements of the algorithm. Section 7 describes the effects of presenting the results of the POV analysis to the broadcaster. And we describe the results of the additional implementation and evaluation experiments. Section 8 describes a discussion and Section 9 summarizes this study.

## 2 USE CASE

The use case which we envision for this research is that a single broadcaster delivers the situation of walking through a tourist spot. The broadcaster will visit a tourist spot and report about the spot to the viewers. The viewers can also request a report to the broadcaster using comments and they will be able to post their impressions of the broadcasting. The equipment used for the broadcast shall be a laptop computer and an omnidirectional camera. The broadcaster must carry a backpack with a camera mounter. Figure 1 shows the broadcaster who carry a backpack. The 360-degree video is centered around the upper part of the broadcaster's back. In 360-degree Internet broadcasting, the broadcaster will have access to more different means of communication than that of conventional broadcasting. For example, he/she can ask to direct attention to the object, or

ask viewers to look for something from their surroundings. The broadcaster's motivation to use it is to make the walk in the tourist spot better. The idea is that the broadcaster can gain a virtual companion from the viewers, even if the broadcaster is traveling alone. In addition, the viewers' motivation to use it is to experience the virtual tourism without time and space constraints. The 360-degree Internet live broadcast provides a highly immersive experience as it allows the viewers to watch the full-sky image in any direction. The viewers can get a real sense of the sights as if they were there. Furthermore, the viewers can enjoy the tourist spot without knowledge about that place by taking advantage of the broadcaster's tour because the broadcaster may have some kind of objective for the tour and act like a tour guide. On the other hand, the problem of the broadcaster's inability to grasp the viewers' POV is synonymous with inability to grasp the companion's gaze. It may prevent smooth communication between the broadcaster and the viewers.

### 3 RELATED WORK

There have been many studies on the role of non-verbal information in communication. In particular, gaze information has been shown to play an important role in communicating mutual intentions. The GAZE Groupware is a study of gaze information in communication [1]. In this study, the non-verbal information of the remote communication in a teleconference system is analyzed. He verified whether natural communication can be performed by conducting a meeting with nonverbal information in a virtual conference room. In addition, he discovered a problem that it is difficult to present gaze information because the space in which the conference participants reside is different in the remote meeting systems. He concludes that it is possible to analyze who talks about what by talking about the gaze directions of the communicatees.

Another study on mutual gaze in remote communication using videoconferencing systems [2] has revealed some interesting findings. The authors argue that the eye contact information of the communicatee is an important factor in the outcome of collaborative work with remote communication. Furthermore, the study also examined the method of presenting gaze information and concluded that the presentation of images including the eyes of the communicatee requires a certain size of images. In 360-degree Internet live broadcasting, the POV is the information that indicates the viewing direction and viewing range of the viewers, and it plays the same role as the gaze in remote communication.

On the analysis of viewers' POV in 360-degree video, a study of Yen-Chen Lin et al. examined on the correction of viewing direction in 360-degree video [3]. In this study, they examined a method of correcting the viewer's direction to the direction of the main story of a 360-degree video. They have implemented and evaluated two patterns of corrections: an automatic correction function and a correction with annotations. The results showed that there were multiple purposes and patterns in the viewer's viewing behavior and

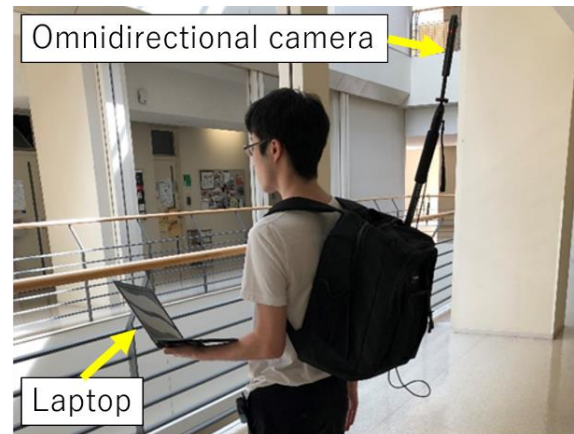


Figure 1: Equipment of the broadcaster.

emphasize the need to analyze the viewer's viewing direction to provide a higher quality viewing experience.

YouTube provides a heat map analysis function for posted 360-degree videos, and the results of the analysis of the entire 360-degree videos are also available [4]. An analysis of the viewer's POV during viewing revealed the characteristics of watching a 360-degree video. The viewer's POV was directed most toward the 90-degree horizontal range centered on the front of the video, where 75% of the playback time was spent. It was also shown that only 20% of the users watched the full 360-degree range, even for the most popular videos.

### 4 HYPOTHESIS TESTING

Based on the results of the related work, we have built three hypotheses concerning the characteristics about the viewers' POV in a 360-degree Internet live broadcasting. The first hypothesis is that "The viewers' POV is concentrated on the direction of the broadcaster's way in mobile environment". In the case of on-demand 360-degree videos, the object matter is displayed in the frontal direction because the video contributor will take or edit the video so that the viewers can fully enjoy the object matter of the video without change of POV. However, the 360-degree Internet live broadcasting is in real time and cannot be edited. Therefore, it is not possible to set an explicit frontal direction to the video, and the viewers may understand that the direction of the broadcaster's way can be the frontal direction. The second hypothesis is that "If the viewers' POV directs at other direction except the direction of the broadcaster's way, the viewing behaviors have meanings and there are some interesting objects for the viewers in the direction". In a 360-degree Internet live broadcasting, viewers can change the POV according to their own interests. As shown in related work[4], viewers are most likely to watch in the frontal direction, and when they watch in directions other than the frontal direction, their interests and concerns are likely to be directed in that direction. The third hypothesis is that "The viewers' POV returns to the direction of the broadcaster's way after the viewers' interests are satisfied". If the viewer's interest is satisfied or the target of interest is no longer visible to the other direction, they no longer need to view to the other direction and return their viewing to the frontal direction.

Summarizing the above hypothesis, in a 360-degree Internet live broadcasting, the viewer's POV is directed in the direction where the broadcaster is going. It changes from the frontal direction to the other direction when a target of interest is found. Thereafter, when the interest is satisfied or the target is no longer visible, the viewer's POV is expected to return to the direction where the broadcaster is going.

We conducted a broadcasting experiment to test these hypotheses. The purpose of the experiment is to test the three hypothesis and collect the data needed to create an analysis algorithm. Six collaborators participated in the experiment, and three collaborators each played the role of a viewer, and the experiment was conducted twice. The role of the broadcaster was played by a same member of the research team. The location of the broadcast is Takamatsu Pond in Morioka City, Iwate Prefecture, which is famous as a place where swans fly and a famous place for cherry blossoms. We made a 30-minute broadcast while moving around the pond, once for right and once for left. We explain to the collaborators that the purpose is to test the operation of a 360-degree Internet live broadcast. The POV was stored in the POV server every 100 milliseconds, and the hypothesis was tested by analyzing the POV log after the broadcast.

Since the POV logs could store 18,000 data per person in a 30-minute broadcast, 1,080,000 data were collected in two broadcasts. The analysis showed that the time when the POV was directed to a 90-degree range centered on the direction of the broadcaster's path was 75.89% in the first experiment and 80.33% in the second experiment. Since the results of the analysis were close to those of the YouTube report, and we decided to proceed with the data analysis based on the collected data. From this point, we followed the YouTube report [4] and called a 90-degree area centered on the broadcaster's direction of walk as the front range. And the ranges other than the front range will be called the other range. Figure 2 shows the equirectangular video and the front range. In addition, we confirmed that Hypothesis 2 was valid because we found several cases in which the target of the viewer's interest was the same as the POV when the POV leave from the front range. We confirmed that the return of the POV to the front range an average of 11.41 seconds after the POV leaves from the front range. This confirms that Hypothesis 3 is also valid. The algorithm acquires the direction of the broadcaster's way as horizontal coordinates from 0 to 360. Then, based on the latest POV data for 10 cases, the algorithm decides which of the four states corresponds to which one of them is applicable, using a conditional expression. However, since the POV data is acquired every 100 milliseconds, immediate state transitions would lead to many false positives. For this reason, we adopt the stacking method for state transitions and do not transition until 10 different states are input. Also, the number of inputs is reset for each of the states that have been entered when the state transition occurs. The algorithm is able to deter the false detection of users viewing near the front range boundary. Due to the nature of the omnidirectional camera, it is not clear how the camera will be installed and fixed, so it is expected that the direction of the broadcaster's way will be different for each broadcast. In

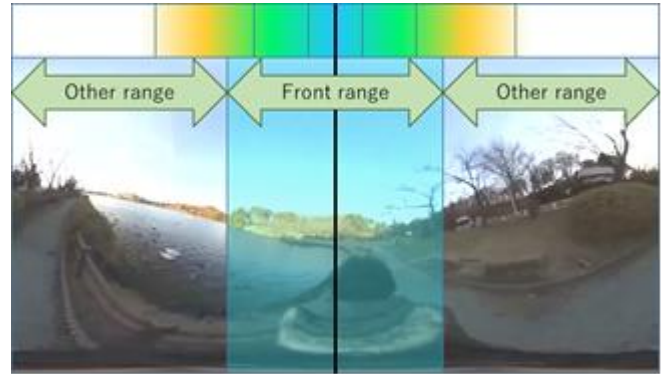


Figure 2: The equirectangular video and the front range.

order to implement the algorithm, you need to acquire and compensate the direction of the broadcaster's way using acceleration sensors.

## 5 ALGORITHM

By testing three hypotheses, we found that viewers changed their POV according to their own interests when the POV was directed to the other range. Therefore, we developed an algorithm to detect POV viewing within other ranges. We determined the classification of the viewer's state and the buffer size to be used in the detection algorithm based on the data used to test the hypothesis. The viewer's state is classified into the following four categories. The state in which the viewer is viewing the front range is called the "normal viewing". The state in which the POV changes from the front range to other range is called "start of other range viewing". The state in which the viewer is continuously viewing the other range is called "other range viewing". The state that returns to the front range is "end of other range viewing". The buffer size for the analysis was set to 10 data of POV. We compared the detection accuracy and immediate response, it was determined that this buffer size was the most appropriate for. Normal viewing is the state which the viewer's POV changes only within the front range, and we are expected to remain in this state for the longest period of time during the broadcast. The algorithm detects and analyzes the POV in the state of other range viewing by triggering the start of other range viewing and end of other range viewing. Figure 3 shows a flowchart of the algorithm we created. Figure 4 shows the variables and conditional expressions used in the flowchart.

We verified the algorithm using the collected data if it correctly detected the POV of other range viewing. As a result, algorithm detected the other range viewing 73 times of at the first broadcast, and detected 95 times at the second broadcast, for a total of 168 times. We compared each results of detection with the recorded video to see what was being viewed. There were 149 cases (88.69%) in which the target object was obvious, and there were 19 cases (11.30%) in which the target object was unclear. Detection results that were unclear on the target object were mainly operation checks and a search of the area. This confirms that the POV pointed the other range was manipulated to view something. We divided 149 data into two groups in terms of whether they can be used for communication, such as whether the objects can be used as topics for broadcasting. There were



76 (51.00%) cases that were judged to be useful for communication by the broadcaster and 73 (48.99%) cases that were judged to be difficult to use. There was no significant difference between the two classifications in terms of whether they can be used for communication.

Those that can be used for communication are called Group A. Those that are difficult to use for communication are called Group B. Those that objects could not be identified are called Group C. Group A's viewing objects were a small shrine built on the bank of a pond and a duck landing on the water that had flown in. Group B's viewing objects were cars and flocks of birds parked around the perimeter of the pond. The data for Group C was a confirmation of operations and a search of the area immediately after the start of the broadcast. After analyzing the data for each group, we found that average time of other range viewing of A, B and C were 9.76, 12.09 and 15.36 seconds respectively. The standard deviations of A, B and C were 7.89, 23.37 and 23.27. However, each of B and C had one data item which duration of other range viewing was more than 100 seconds and they might contain extreme outliers. When these outliers were removed, we found that time of other range viewing of A, B and C were 9.76, 9.73 and 10.44 seconds, respectively. The standard deviations of A, B and C were 7.89, 12.13 and 10.54. From additional interviews, we found that the viewers interrupted and left the POV operation to enter comments.

## 6 IMPROVEMENTS

The algorithm detected the POV of the viewers directed to other ranges. The significant data detected by the algorithm were 149 of 168 (88.69%) Data. However, the number of data classified as useful for communication was only 51.00% (76/149). In Group B, there were many detections that had already noticed by the broadcaster because the viewers were viewing near the boundary of the range. Therefore, it is necessary to re-examine the boundaries of the front range setting in this study. The POV is information that indicates in which direction the center of the viewing image is pointed. Therefore, when the object of viewing is at the edge of the screen, the POV is directed to the other range but viewers may view front range. In the 360-degree Internet live broadcasting system, the camera angle of view in three.js is set at 35 degrees. To completely hide the front range, it is necessary to point the POV at a range of  $\pm 62.5$  degrees or more. We decided to redefine the boundaries of the algorithm to  $\pm 60$  degrees to analyze the data again.

As a result, new algorithm detected other range viewing 127 times, and we found that Group A and Group B were 114(89.76%). New algorithm detected Group C 13(10.23%). Among the 114 data items, data classified as A and B were 70(61.40%) and 44(38.59%). By improving the algorithm, we were able to prevent false positives and improve the percentage of group A by about 10%. By extending the front range, 41 viewings near the boundary of the front range were excluded from the detection of other range viewing. 35 data that were excluded from the detection were classified as Group B or Group C. We were able to exclude from the detection those that were close to the boundary line, which

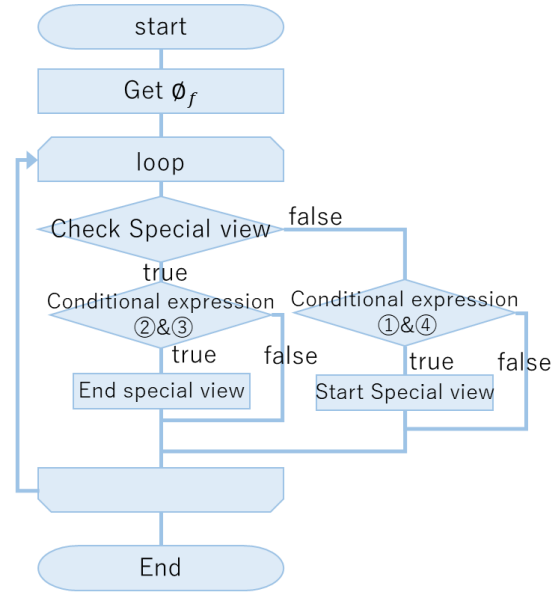


Figure 3: Flowchart of the algorithm

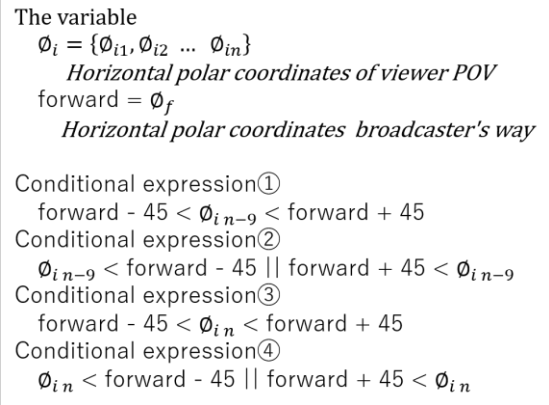


Figure 4: The variables and conditional of flowchart.

were less topical, such as people looking at the pond, people passing by, and trees growing beside the sidewalk. On the other hand, 6 data classified in group A were excluded from the detection. Excluded from the detection were the pond management office building, a swan boat covered with a blue sheet, a passerby with a camera, and a duck landing. The reason for these being out of detection is that the broadcaster was standing still and could watch without having to move the POV significantly. Therefore, POV was included in the extended front range. The algorithm may be difficult to detect other range viewing when the broadcaster is at a standing still.

## 7 EVALUATION

To present the results of the algorithm's analysis, we implemented an additional POV indication function. We superimpose different colored circles on the equirectangular video of the broadcaster's user interface in order to show the viewers' other range viewing. The colors are different for each user, and the center of the circle shows the POV. Figure 5 shows the superimposed display of the POV. The diameter of the circle matches the horizontal 25 degrees of the video. The viewers are viewing a 75-degree horizontal



angle of view through the viewer's user interface, but it is impossible to view the entire 75-degree range of the video. The display used in this experiment is 21.5 inches with a resolution of 1920 x 1080. It is full HD quality, and the video displayed on the viewer client is 14.28 cm x 10.71 cm. It is said that the field of view where a human being can process information accurately without moving his or her eyes and head is 5 degrees or less in the horizontal direction. If the distance between the viewer's eyes and the display is 60cm, the range is 5.24 cm or less in the horizontal direction. The viewer can watch only about one-third of the video in the horizontal direction. Therefore, the indication of the POV is based on a range of 25 degrees in the horizontal direction, which is one-third of the camera angle of view. We have decided to present them. This criterion is used as a test for this evaluation.

An experiment was conducted to present the results of viewers' POV analysis to the broadcaster using an improved algorithm. The broadcast experiment was conducted twice. The broadcast route, broadcast time and equipment were the same as in the hypothesis test. In this experiment, there were eight experiment participants, because the role of the broadcaster was also played by the experiment participants. The broadcaster was informed of the broadcast procedure and that the viewers' POV would be displayed on a laptop computer. They were also instructed to stop at the edge of the sidewalk when checking display of the POV. No instructions or physical restrictions are placed on the content of the broadcast. We only told the viewers that we would be conducting a 360-degree Internet live broadcasting test. The expected effects of the experiment are the improvement of understandability of the viewers' broadcasting needs and events which could not be noticed by the conventional broadcasting system. This is because the algorithm only extracts the necessary POV, which reduces the number of POVs that the broadcaster needs to check and makes it easier for the viewer's interest to be identified. After the experiment, we extracted data of other range viewing from the POV log and presented the broadcaster together with the recorded video for interview. For each of the other range of viewing, we asked whether they noticed the POV indication, whether they grasp the object from the POV indication, whether the object could be used as a topic, and the reason why it was topical. Additional questions were asked in an open-ended format if we had interested in the answers.

From the experimental results, the detecting other range viewing by the improved algorithm were Group A and Group B were 35 and 48. Table 1 shows the results of the interviews with the broadcasters after the experiment if there were the topicality and the reasons for the responses. In some cases of other range viewing, the viewers watched same objects at the same time. In the case of presence of topicality, it was the pond or birds whose state was frequently changed by the flight of wild birds and the movement of carp. For example, it shows a main attraction of the pond such as a rowing swan boat covered with blue sheets. In the case of absence of topicality, objects whose state was not frequently changed, such as street trees and buildings, or objects which do not particularly catch your

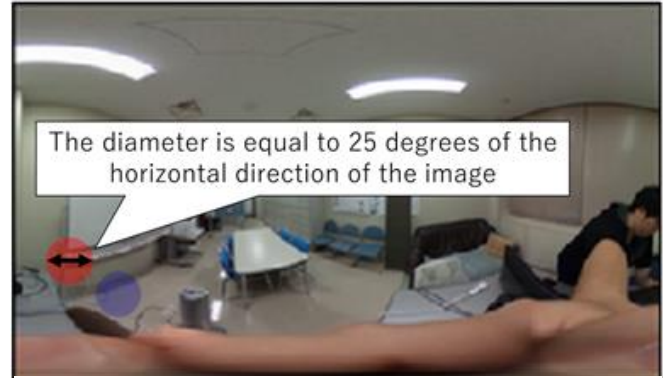


Table 1: The percentage results of the topical and notice.

broadcast	Topical	Notice	Percentage
1st broadcast	Yes : 20 (57.14%)	14 (40.0%)	62.86% (22/35)
	No : 15 (42.86%)	8 (22.86%)	
2nd broadcast	Yes : 39 (84.78%)	22 (47.83%)	54.35% (25/46)
	No : 7 (15.22%)	3 (6.52%)	

eye, such as passersby, were mentioned. We thought the broadcaster have to stop once to check the POV indication and might not notice the POV indication because he/she cannot watch the screen of the laptop during the walking. However, from the interviews, we found that the broadcasters frequently stopped to check their comments and were able to notice the POV indication. The number of times they noticed the POV indication was 62.86% (22/35) for the first broadcast and 54.35% (25/46) for the second broadcast.

## 8 DISCUSSION

The reasons for the presence or absence of topicality can be divided into three categories. The first reason is that changes in the condition of the object or the state of the object are topical. The broadcaster can keep talking about changeable objects such as wild birds and frozen ponds. The second reason is that it enables the broadcaster to confirm the interest of the viewers, which was not confirmed by the comments. The only means of communication from the viewer to the broadcaster is performed through text-based comments, and the broadcaster cannot understand interest of the viewers unless the viewers send comments about their interest on their own initiative. In the experiment, the broadcaster noticed that the viewers were interested in boats that were not mentioned in the comments, and could use them as topics of conversation. We also found that viewers may be interested in the same subject matter. The third reason is that the broadcaster can confirm viewers' interests about what the broadcaster said. In the experiment, although the broadcaster made frequent calls to viewers, the response comments was not that great. However, the broadcaster could confirm the viewers' interests of an object which the broadcaster talked about by checking the POV indication directed at the object of the topic.

From these results, by presenting the POV of other range viewing, the broadcaster can not only adapt the topic to the viewers' interests, but also get the response to the broadcaster's statement from the POV indication. Furthermore, three criteria potentially can be used to analyze the topicality of the POV. In half of the cases, we found that the broadcaster could not be aware of the displays of superimposing the POV on the broadcaster's user interface. It is necessary to consider a method of presenting the POV so that the broadcaster can check the POV indication even while walking. As a concrete method, we can use senses other than vision, such as sound and vibration. It is also necessary to consider a method for automatic identification of topicality because some POV are presented without topicality.

## 9 CONCLUSION

In this study, we investigated and improved a POV analysis algorithm that uses the viewers' viewing behavioral characteristics in a 360-degree Internet live broadcasting. Behavioral characteristics were validated and we implemented an algorithm that uses viewers' behavioral characteristics. We conducted initial evaluation of the algorithm was conducted, and we could detect the significant POV changes. Further improvement of the algorithm was investigated to increase the percentage of beneficial data which can be utilized for broadcasting, we found that significant POV with 89.76% accuracy. As a result of the experiment for the impact of presenting the analysis results, we found that POV indication made it easy for the broadcaster to understand interests of the viewers. In addition, we found that the inclusion of non-topical other range of viewing in the POV indication will not have a significant negative impact on communication and broadcasting.

## ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP20K11794.

## REFERENCES

- [1] Roel Vertegaal, "The GAZE groupware system: mediating joint attention in multiparty communication and collaboration", CHI '99 Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 294-301 (1999).
- [2] David M. Grayson, Andrew F. Monk, "Are you looking at me? Eye contact and desktop video conferencing", ACM Transactions on Computer-Human Interaction (TOCHI) Volume 10 Issue 3, September 2003, pp. 221-243 (2003).
- [3] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, Min Sun, "Tell Me Where to Look: Investigating Ways for Assisting Focus in 360° Video", CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 2535-2545(2017).

- [4] YouTube Creator Blog: Hot and Cold: Heatmaps in VR, <https://youtube-creators.googleblog.com/2017/06/hot-and-cold-heatmaps-in-vr.html>

# A Privacy-oriented Video Distribution Platform for Public Camera Systems

Satoru Matsumoto<sup>\*</sup> Tomoki Yoshihisa<sup>\*</sup> Tomoya Kawakami<sup>\*\*\*</sup> Yuuichi Teranishi<sup>\*\*\*</sup>

<sup>\*</sup>Cybermediacenter, Osaka University, Japan

<sup>\*\*</sup>Graduate School of Engineering, University of Fukui, Japan

<sup>\*\*\*</sup>National Institute of Information and Communications Technology, Japan  
{smatsumoto, yoshihisa}@cmc.osaka-u.ac.jp

**Abstract** - Recently, public cameras of which the main purpose is crime prevention are installed at various places. The videos shot by public cameras are also useful for people other than the staff and can be helpful for various purposes. We call this concept, i.e., general people can check the videos shot by public cameras, the Next-Generation Public Camera Systems (NGPCS). In NGPCS, general people (the viewers) can get the information of the people shot by the cameras. This information includes the privacy information about the locations or the situations of the shot people and thus has a large possibility to cause public concerns. However, conventional public camera systems are managed by reliable security companies, and thus do not protect privacies of the shot people in many cases. In this paper, therefore, we propose a privacy-oriented video distribution platform for NGPCS. Our proposed privacy-oriented video distribution platform for NGPCS is required faster human detection, flexible policy description, and faster video processing. To realize them on the platform, we propose three technologies. In this paper, to check the effectiveness of our approach, we designed and implemented our proposed platform. Also, we measured and compared the video processing time under the technology with that under a conventional technology.

**Keywords:** Cloud Computing; Edge Computing

## 1. Introduction

Recently, public cameras of which the main purpose is crime prevention are installed at various places. Security companies manage them, and the staffs check the videos shot by them. The videos shot by public cameras are also useful for people other than the staff and can be helpful for various purposes such as the safety checks of children or finding pets ran away. We call this concept, i.e., general people can check the videos shot by public cameras, the Next-Generation Public Camera Systems (NGPCS). In NGPCS, general people (the viewers) can get information about the locations or the situations of the people shot by the cameras. Here, we assume live videos so that they can get real-time information. This information includes the privacy information about the locations or the situations of the shot people and thus have a large possibility to cause public concerns. For example, a thief checks the situation in a house by an NGPCS and finds a good timing for breaking into the house. In another example, a man finds the location of his friend and exposures the friend's activity. Therefore, NGPCS requires privacy protections for using it safer.

Since it is difficult to identify the owner of the object to be shot, frameworks for properly designing and implementing a privacy protection mechanism has not yet been established, and it is impossible to realize a public camera system that protects privacy. Many web sites show images of public cameras, but they do not protect privacy and are open to the public. Therefore, these web sites can cause public problems. Conventional public camera systems are managed by reliable security companies and, thus, do not protect the shot people's privacy in many cases. Even the case that the privacy protection is considered, only a signboard notifies the camera shot areas and the pedestrians should avoid the area not to be shot by the camera.

Hence, in this paper, we propose a privacy-oriented video distribution platform for NGPCS. In the proposed platform, a video processing server detects objects in the recorded videos. The server adds some video effects to protect the privacy of the shot objects based on the privacy protection rules. The privacy protection rules are written by the shot objects, and if there are no rules for the shot objects, the server uses the safest default rules. The rules are written in the form of If-Then rules. The novelties of this research are the concept of NGPCS and the focus on the privacy issue on NGPCS. The contributions of this research are: 1) the design of the privacy-oriented video distribution platform 2) a practical development of the platform 3) the performance evaluation of our developed privacy-oriented video distribution platform. One of the simple processes of our proposed platform is "finding person and then hide the person's face," and this seems to be a kind of straightforward approach. However, this is just one case and this research contains some novelties and contributions as written above.

In Section 2, we introduce related work. In Section 3, we explain our proposed platform and evaluate it in Section 4. Finally, we conclude the paper in Section 5.

## 2. Related Work

In [1,2], we have proposed the concept of the different world broadcasting, in which various video effects give the feeling that the broadcasters are in another world. In [3], we have proposed a trust-oriented video distribution system, which considers the trust between the broadcasters and the viewers. In these researches, each camera is managed separately by the broadcasters. On the other hand, in the NGPCS proposed in this paper, each camera is often managed by a security company and can construct the platform systematically considering the processing loads.

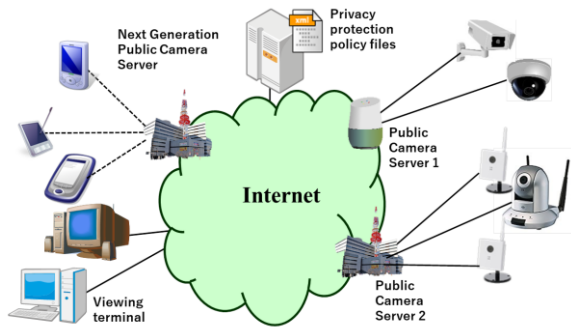


Figure 1: Assumed system

There are many kinds of research on reducing the processing time of Cascade recognition ([4]). In [5], a method to distribute the computational for with the process to extract objects from video shot by public cameras to multiple computers is posed. In the proposed method, the contents of the video processes are fixed. However, they change depending on the object to be shot. This research differs in that the images are processed based on the privacy protection policies to be shot.

In [6], an image process description method was proposed to visually recognize various image processing for images taken with public cameras. By using this method, it is possible to describe the processing method flexibly for each object to be shot. However, we use the rule style description in the NGPCS because it is difficult to parse the description in the method.

Some researches focus on the privacy issue for the surveillance camera systems and propose systems to hide personal information such as faces [7-9]. Different from these researches, we focus on the privacy issue on NGPCS. Moreover, we design, develop, and evaluate the our proposed platform in this paper.

In [10], the authors proposed a method to determine the locations to install public camera considering the covered area for their shoot. Even in the NGPCS, the wider the shooting area is, the more various applications can be used, but the research target is different from this in the point that we focus on the privacies of the shot objects.

Focusing on the distributed fault-tolerant processing method, Yang et al. proposed a distributed image retrieval method [11]. They combine cloud storage technology, data encryption, and data acquisition technology to enable efficient integration and management of public cameras. However, similar to the previous research, they do not consider privacies of the shot objects.

### 3. Privacy-oriented Video Distribution Platform

#### 3.1 Assumed System

Our assumed system is shown in Figure 1. The public camera server (PCS) can communicate with the next-generation public camera server (NGPCS) via the internet and can send and receive video data shot by the public camera and privacy protection policy files. The privacy protection policy files inscribe the privacy protection policy of the owner of the object to be shot. We assume that the

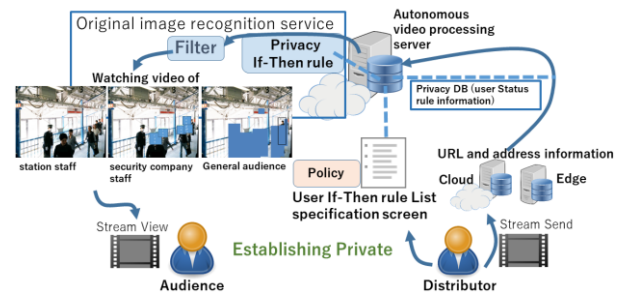


Figure 2: The architecture of our proposed privacy-oriented video distribution platform

owners log in to the NGPCS in advance and write it. If he/she does not write it, the platform executes the video processes to hide the shot objects, which is considered to be the safest. Also, in this research, we utilize the concept of edge computing in that the platform exploits the computers on the edge of the network, which is attracting attention in recent years. By using this concept, ECSs (edge camera servers) cooperate to transmit and receive video data and identify the same object to be shot.

#### 3.2 Requirements for the Privacy-oriented Video Distribution Platform

There are 3 main challenges for realizing the privacy-oriented video distribution platform.

- Faster human detection: longer processing time for detecting humans shot by the cameras causes a longer delay in checking the videos.
- Flexible policy description: we can use various approaches to protect privacies such as hiding the human faces by virtual masks, blurring the human bodies, and so on. The preferable approaches depend on the viewers and shot humans. Therefore, it is required for our proposed platform to be able to describe the policies flexibly and find the policy to be applied faster.
- Faster video processing: our proposed platform adds some video effects to the shot videos to protect the privacies. Longer processing time for this causes a longer delay in checking the videos.

#### 3.3 Faster Human Detection

In our proposed platform, the objects to be shot is identified to add some video effects based on the privacy protection policy of the owners of the shot objects. If the NGPCS executes the processes of identifying the object and the execution of adding video effects, the processing load will be concentrated on the NGPCS, resulting in long processing time and a long delay for watching the video. Therefore, in this research, we propose a technology called *edge camera server cooperation* utilizing the concept of edge computing that executes the processes at the edge of the network, which has attracted attention in recent years.

In the edge camera server cooperation technique, the ECS cooperate to execute video processes with the cloud servers and the local computers to reduce the processing loads. For example, after the kindergarten child is shot from the front by the ECS 1. An image of the back style is also shot by the same server. Even if the ECS 2 shoots only the back of a

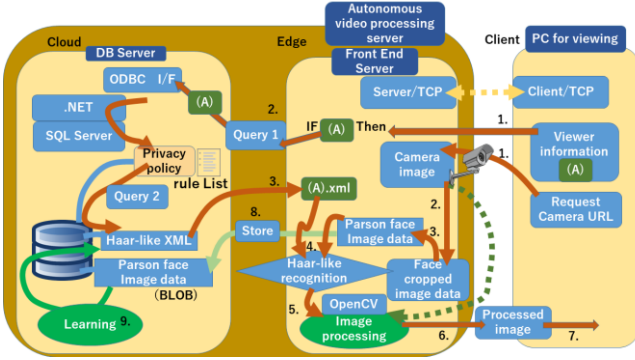


Figure 3: The overall system behavior for the experimental evaluation

kindergarten child in a store, the kindergarten child can be identified by comparing the image with the back image shot by the ECS 1.

### 3.4 Flexible Policy Description

In our proposed platform, the privacy protection policy of the owner of the shot objects is described in the privacy protection policy. The privacy protection policy files inscribe a video processing method such as hiding the object to be shot by a rectangle, applying a mosaic, or adding an annotation according to the features of the viewer (occupation, age, sex, etc.). For example, as shown in the left-upper side in Figure 2. If the viewer is the person to be shot, other people are covered with a rectangle, and if it is a staff member, the face is hidden by the customer number. When a complicated process related to the privacy protection policy is described by a program, there is a problem that the description becomes complicated and difficult. Therefore, in this research, we propose a technique called the *rule-based privacy protection policy description*.

In the rule-based privacy protection policy description, the privacy protection policy is described by rules in the style of If-Then rules. The privacy protection policy can be described by combining simple rules, and the privacy protection policy can be flexibly described. An example of the If-Then rule follows. If: a viewer is a general person, and the object to be photographed is a person. Then: the whole body is covered by a rectangle.

### 3.5 Faster Video Processing

In our proposed platform, the video shot by the public cameras is processed based on the privacy protection policy file to generate the video that the viewer watches. When the video is generated for each viewer, the processing load of the ECS related to the video generation becomes large, which causes a problem that it takes time to generate the video. Therefore, in this research, we propose a *local image processing recognition* technology.

In this technology, the privacy protection is analyzed in advance, and find the video processes that the local machine can execute. Since the processing loads for the processing servers are reduced, faster video processing can be conducted. For example, the local machine crops the face region for blurring on it before sending it to the processing servers.

Table 1: Components and performance

Machines	Metric	Performance
The cloud server as NGPCS.	VM Plan	• Standard B2s( 2vcpu 4 Gib memory)
	OS	• Windows 10 Pro
	CPU	Intel(R) Xeon(R) platinum 8171M CPU @2.60Ghz
	Memory	RAM 4.00GB
The edge server as ECS	OS	Windows10 Pro
	CPU	Intel(R) Core(TM) i7-7660U CPU @2.50GHz
	Memory	RAM 8.00GB
the local computer as the viewer's PC	OS	Windows10 Pro
	CPU	Intel(R)Core(TM)i5-6300U CPU @2.40GHz
	Memory	RAM 8.00GB

## 4. Experimental Evaluation

### 4.1 Experimental Environment

The authors constructed an experimental environment for our proposed privacy-oriented video distribution platform for NGPCS. For the experimental environment, in this paper, we use some virtual machines to simulate the cloud server and the edge server. The database for storing the privacy protection policy files and others is also built in the virtual machines. We constructed them on the cloud computing resource provided by Windows Azure. The image of the architecture of our implemented environment is shown in Figure 3. The detailed processing flow is explained in the next subsection.

Table 1 shows the detail of the OS, CPU, and main memory of the servers used for processing in the evaluation. The cloud server is a virtual machine built on Microsoft Azure service, and the VM plan is an inexpensive plan for databases, which are servers without a high-speed GPU, but better than the local computer as shown in Table 1. A notebook PC is used for the watching machine.

The image usage policy is saved as a rule list in the Database, and the XML for face recognition corresponding to each rule is also saved to the database. After the If-Then rule is determined, the face recognition process is executed. SPC consists of two servers. The image processing server is a pre-processing server, and the database server is a post-processing server. We submit queries to the SQL database for If-Then rules and retrieve the corresponding XML for authentication.

The design of each of the 3 main tables used in the database follows.

(a) The columns are rule number, XML data file name, and XML data set the table used for image recognition. This table is used to get the Then part of the If-Then privacy protection rules.

(b) The columns are the pairs of personal identification numbers and applicable rule numbers. The rule number is the main candidate key and duplication is allowed. This makes it possible to deal with complicated rules having multiple rules.

(c) The columns are personal identification number, applicable rule number, image RAW binary data set of the person. Each record in this table is a transaction record.



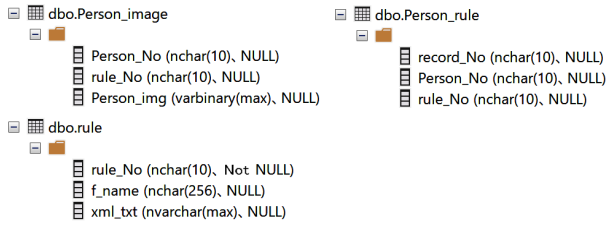


Figure 4: Database structure

The data of the faces that do not match in the face recognition is temporarily saved for later classification.

These 3 tables are shown in Figure 4.

We are also designing an application that registers a new XML file in the database. Figure 5. Shows the XML rule data upload interface.

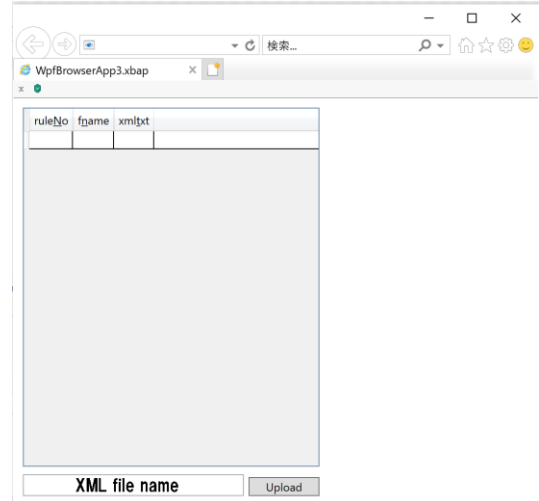


Figure 5: XML rule data upload interface

## 4.2 Experimental Configuration

### 4.2.1 Configuration for Checking Faster Human Detection

By the use of edge camera cooperation technique, the human detection processes, and servers that accept requests for watching public cameras are allocated to the ECS provided by cloud computing resources. To check our approach for the faster human detection, we use the cloud server and the edge server and measure the processing time and the frame interval. The ECS uses OpenCV4.3 to perform face authentication by using Haar-Like feature extraction, detects the person, executes adding masks, and sends the image to the requested client machine.

The process flow from requesting watching the video from the client machine to watch the video on the client machine follows as shown in Figure 3.

1. The user makes a request from the client machine to the public camera using the URL and sends the viewer's identification number simultaneously.

2. Based on the request received by the ECS, it submits the queries to the database to get the privacy protection rules and the XML file for face authentication. At the same time, the requested PCS sends the video data to the ECS.

3. The ECS executes the face recognition process and adds some video effects based on the privacy protection rules written in an XML file.

4. If it is a newly registered user, to perform accurate detection, the ECS gets his/her face images and updates Harr-Like feature data.

5. Utilizes OpenCV to process video according to the then part of the If-Then rule and sends the video to the client machine for watching the video.

8. To add his/her face images as an XML data for OpenCV, ECS crops the face region as a Binary Large Object (BLOB) data and create it. After that, the ECS registers it into the database together with the personal identification number.

9. The PCS feedbacks the Haar-Like cascade data for learning results in the XML file.

### 4.2.2 Configuration for Checking Flexible Policy Description

To check our approach for the flexible policy description, we check that our proposed platform surely works If-then rules. A list of rules for authentication, XML data for Haar-Like feature extraction, and person image data for authentication newly added by If-Then rule to the list. These data are expressed by a BLOB. To learn the condition and the image recognition condition characterized by the If-Then rule, BLOB data is stored in the database until the learning is completed. This is because there is a risk of privacy leakage when the raw person image data is stored on the server for a certain period.

### 4.2.3 For Checking Faster Video Processing

To realize faster video processes, in this paper, we tried to speed up by clipping only the face part of the human image, and sending it to the cloud server, and execute image recognition only to the clipped region on the cloud server.

## 4.3 Evaluation Method

We evaluated the performance of our proposed platform. To show the performance numerically, we measured the processing time for the human (face) detection and the processing time for the video processing (blurring). Regarding about the flexible policy description, we checked that our proposed policy description, If-Then rules worked properly.

Using the Harr-Like cascade classifier, we measure the image processing time for each frame. The image processing time includes the time to cut out the recognized face, the time needed to blur the image and show it on the Local computer. After the feature classification recognition process in the Local computer, only the cut-out part is sent to the cloud server and the edge server. These servers return the processed image to the Local computer. The local computer combines the returned image with the original image and shows it on the screen. Although our proposed platform is designed for executing various processes, we focus on only the blurring effect to make the results consistent.

The video for the experiment is a video for an orchestra concert and includes 38 people for playing musical

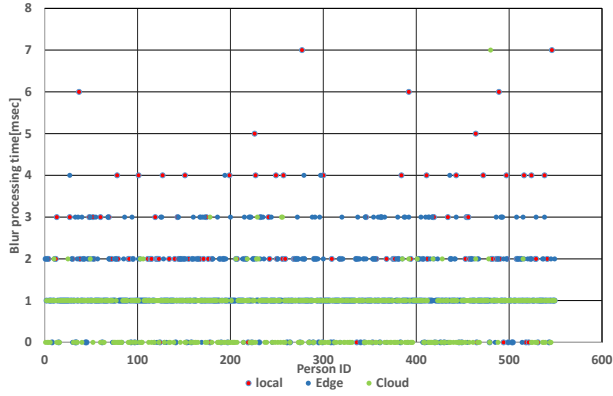


Figure 6: The processing time to blur the face area

instruments. We have already measured the turnaround time in [7]. The turnaround time is the time to start sending the data on the local computer to finish receiving the processed data on the local computer. As a reference, the turnaround time was 16.28 [msec.] Therefore, in this paper, we measure the time needed to add video effects and the time needed for the image processing. These are mainly related to the computational servers and suitable for checking the effectiveness of our approach. Since the writing of Blob image data is post-processing, we do not measure it here.

## 4.5 Results

In the evaluation method of 4.3, the processing time for each frame was measured separately for the case where all the processing was performed only by the Local computer and the case where the processing was distributed to the Cloud server and the edge server.

### 4.5.1 Processing Time for Blurring

Figure 6 shows the processing time to blur the face area. The number of the viewers is 3. In the figure, the local means that the processing time when the process to blur the video image is executed in the local computers. The Edge means that the processing time when the process is executed in the edge servers, and the Cloud means executed in the cloud servers. Since the number of the clients is 3, we used the notation 3. The horizontal axis is the person ID. The person ID is given to each detected person sequentially. The vertical axis is the time to blur the face area for each person.

From this figure, we can see that the processing time under Cloud is the shortest in many cases. This is simply because the processing power of the cloud servers is high. Since the processing power for the edge machine and the local servers are almost the same, their processing time is approximately equivalent in many cases.

However, the processing time is sometimes very long since the local servers are used for other purposes such as running other applications, and so on.

### 4.5.2 Frame Interval

Figure 7 shows the frame interval. The frame interval is the time to recognize the face area, blur the face area. The Cloud/Edge Local means the frame interval when the local computer determines the face region, and the PCS executes the subsequent blurring processing is executed. The local

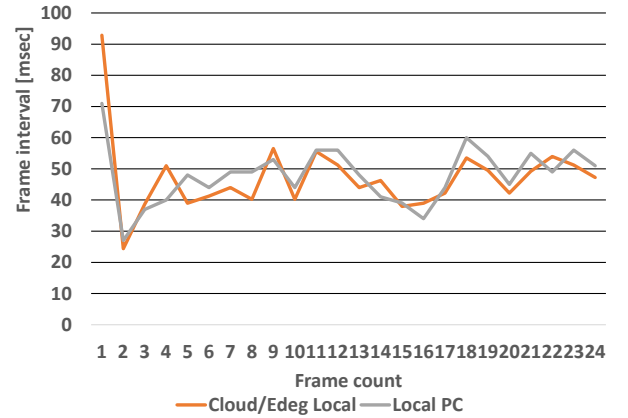


Figure 7: Frame interval under the cloud/edge environment or the local environment

computer means the frame interval when the local computer executes all the above processes. From this figure, we can see that the frame interval is almost the same between the Cloud/Edge Local case and the local computer case. This is because the time for blurring the face region is reduced by using the cloud servers. But, using cloud servers require communication time. In this case, therefore, the frame interval is not reduced largely by using the cloud servers, and the overall processing times for each case are equivalent.

## 4.7 Discussion

As a problem with image processing, when the frontal face is partially hidden by a shield such as a thin rod or a lattice, face recognition fails, and some faces were not detected cannot be executed. In the video for the experiment, an average of 25.6 persons were recognized by their faces in the 67 frames measured with the similarity parameter of 1.08 for the Harr-Like extractor. An average of about 13 people remained unprocessed within one frame. This is because there are obstacles between them and the camera. It is difficult even for our proposed platform to detect faces in such cases.

## 5 Conclusion

In this research, we proposed the Next-Generation Public Camera Systems (NGPCS). In NGPCS, general people can check the videos shot by public cameras. To protect the privacy of the shot people, we proposed a privacy-oriented video distribution platform for NGPCS. Our proposed platform was designed to give faster human detection, a flexible policy description, and faster video processing. To realize them, we proposed three techniques and checked their functions. Our experimental evaluation revealed that our proposed platform gives a faster human detection by using the edge servers, that realizes flexible policy description by the use of If-Then rules, and that reduces the video processing time by cropping the face region of the client machines.

In the future, we will evaluate the performance of our implemented system under the situation that there are more PCSs and more If-Then rules.

## ACKNOWLEDGEMENT

This research was supported by a Grants-in-Aid for Scientific Research(C) numbered JP20K11829 and by G-7 Scholarship Foundation.

## REFERENCES

- [1] S. Matsumoto, Y. Ishi, T. Yoshihisa, T. Kawakami, and Y. Teranishi, "A Distributed Internet Live Broadcasting System Enhanced by Cloud Computing Services," *International Journal of Informatics Society (IJIS)*, Vol. 10, No. 1, pp. 21-29 (2018).
- [2] S. Matsumoto, T. Yoshihisa, T. Kawakami, and Y. Teranishi, "A Design of Hierarchical ECA Rules for Distributed Multi-Viewpoint Internet Live Broadcasting Systems," in *Proc. of the International Workshop on Streaming Media Delivery and Management Systems (SMDMS2018)*, pp. 340-347 (2018).
- [3] T. Yoshihisa, S. Matsumoto, T. Kawakami, and Y. Teranishi, "A Video Processing System to Stabilize Frame Rates on Trust-Oriented Internet Live Video Distributions," in *Proc. of the International Workshop on Informatics (IWIN 2019)*, pp. 31-38 (2019).
- [4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 511-518 (2001).
- [5] M. Saini, X. Wang, and M. Kankanhalli, "Adaptive Workload Equalization in Multi-Camera Surveillance Systems," *IEEE Trans. on Multimedia*, pp. 555-562 (2012).
- [6] R. Cucchiara, "Multimedia Surveillance Systems, in *Proc. ACM Video Surveillance & Sensor Networks (VSSN)*, pp. 3-10 (2005).
- [7] I. Mitsugami, M.. Mukunoki, Y. Kawanishi, H. Hattori H., and M. Minoh, "Privacy-Protected Camera for the Sensing Web," *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 622-631 (2010).
- [8] J. Paruchuri, S. Cheung, and M. Hail, "Video Data Hiding for Managing Privacy Information in Surveillance Systems," *Journal on Information Security*, Vol. 2009, Article number 236139 (2009).
- [9] K. Fujii, K. Nakamura, N. Nitta, N. Babaguchi, "A Framework of Privacy-Preserving Image Recognition for Image-Based Information Services," *International Conference on Multimedia Modeling*, pp.40-52 (2017).
- [10] H. Huang, C. Ni, X. Ban, J. Gao, A.T. Schneider, and S. Lin, "Connected Wireless Camera Network Deployment with Visibility Coverage," *IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1204-1212 (2014).
- [11] J. Yang, B. Jiang, and H. Song, "A Distributed Image-Retrieval Method in Multi-Camera System of Smart City Based on Cloud Computing," *Future Generation Computer Systems*, Vol. 81, pp. 244-251 (2018).



Keynote Speech:  
Dr. Rieko Yamamoto  
( Fellow, Fujitsu  
Laboratories Ltd. )



# Digital Transformation with Trust - Software Development Issues -

Sep. 10, 2020

Fujitsu Laboratories Ltd.  
Fellow  
Rieko Yamamoto

Copyright 2020 FUJITSU LABORATORIES LIMITED

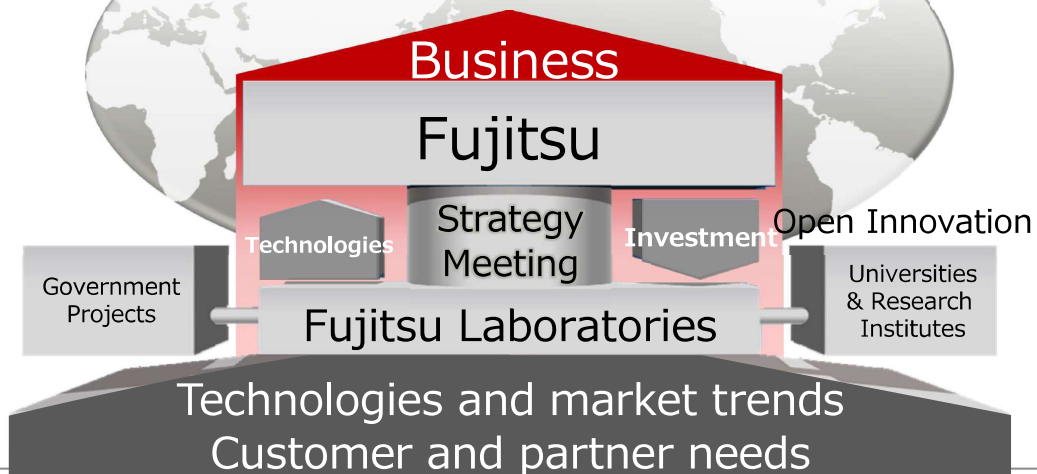
## Scenario

- DX(Digital Transformation) and its Success Cases
- Software Issues for Achieving DX
- Key Areas of Trust in DX Systems Development
  - Examples
    - AI Technology to Automatically Recommend Software Repairs
    - Automatically Adjusting the MSA Environment
    - New Quality Models and Characteristics
    - Existing Systems Analysis Technology
- Improving Experience Technologies

## Introduction to Fujitsu Laboratories: R&D Scheme



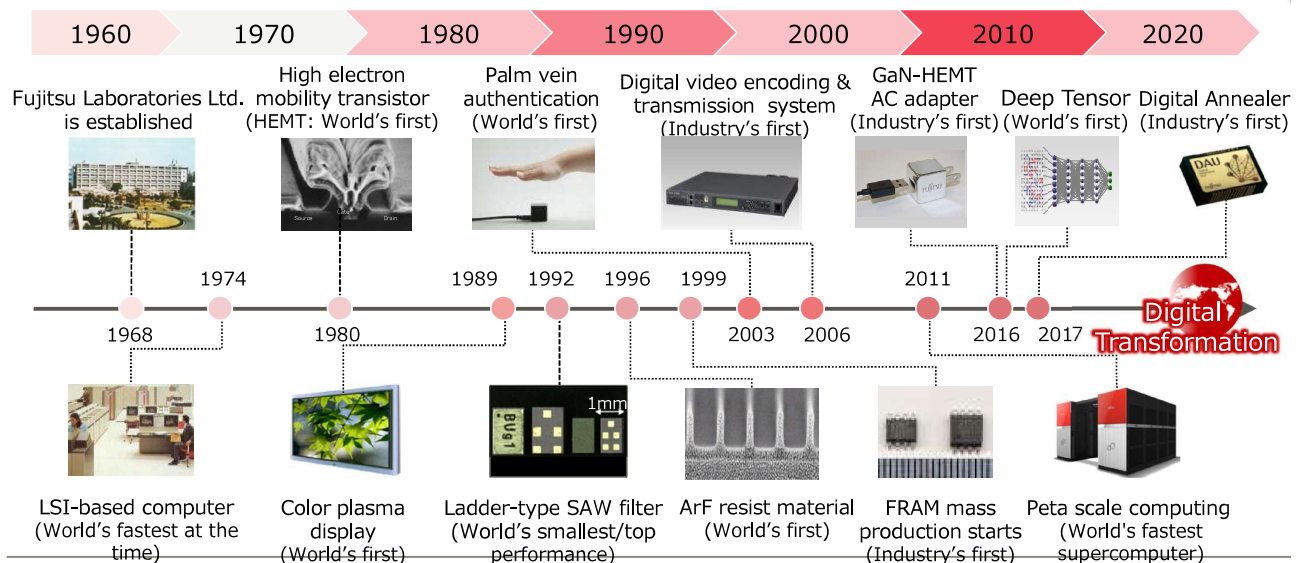
Creation of value, Development of new markets,  
Contribution to main business through Digital Trust and Co-Creation



2

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Fujitsu Laboratories' Innovation



3

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Digital Transformation (DX)



### ■ 3 steps toward DX

1. Digitization: The conversion of analog data to digital data
  - Ex. Digital books, Credit card, Digital Camera, etc.
2. Digitalization: The use of digital technologies and digitized data to change business process more efficiency
  - Ex. Online classes, Internet banking, Online medical examination, etc.
3. **Digital Transformation**: The use of digital technologies to change the business models, social activities and services
  - Ex. Business matching services, Sharing services, etc.

### ■ DX success cases

- **Uber (Sharing services)**, Citigroup (Banking, B2C), etc.

4

Copyright 2020 FUJITSU LABORATORIES LIMITED

## DX Situation in Japan



### ■ METI (Ministry of Economy, Trade and Industry) released a report on Digital Transformation (Sep. 2018). "2025 Digital Cliff"

[https://www.meti.go.jp/english/press/2018/pdf/0907\\_004a.pdf](https://www.meti.go.jp/english/press/2018/pdf/0907_004a.pdf)

Many business owners seem to understand the importance of a digital transformation (DX) as a means of creating new business models or modifying existing ones by taking advantage of new digital technologies in order to further grow their business and enhance their competitiveness. However:

- They face difficulties in data utilization across internal departments due to operational department-based construction and excessive customization of existing systems, causing the systems to be overly complex and closed;
- They face challenges in approaches to carrying out DX, since employees affected by DX often reject such business reforms regardless of business owners' decisions to undergo DX, despite the fact that companies that do not solve the challenges mentioned above in the existing systems to allow for better data utilization, or review the entire business systems are less likely to succeed in the long term.

→ If companies cannot overcome these challenges, Japan may suffer an economic loss of up to 12 trillion yen per year after 2025, three times larger than the current loss ("2025 Digital Cliff") due to the failure to achieve DX.

### ■ DX Promotion Index, Jul. 2018

### ■ DX Promotion Guidelines, Dec. 2018

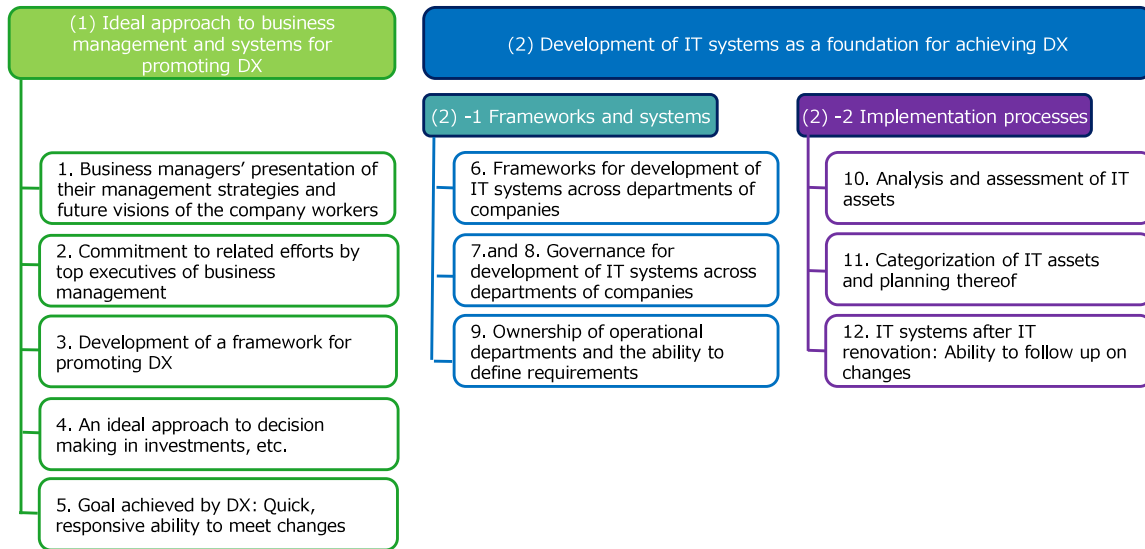
5

Copyright 2020 FUJITSU LABORATORIES LIMITED

# METI DX Promotion Guidelines



[https://www.meti.go.jp/english/press/2018/1212\\_003.html](https://www.meti.go.jp/english/press/2018/1212_003.html)



6

Copyright 2020 FUJITSU LABORATORIES LIMITED

## DX Success Case: Uber



### ■ Uber continuously increases the value of services

- Rapidly expanding new services globally while ensuring business continuity (5.5 million rides per day)

Name	Service contents	Launch / Area
Uber	-	May 2010 / San Francisco
UberX	Four-seater passenger car	Jul. 2012 / San Francisco
UberXL	Six-seater passenger car	May 2014 / San Francisco
UberPOOL	Carpool service	Sep. 2014 / San Francisco
UberACCESS	Wheelchair-friendly car	Dec. 2015 / DC
UberAUTO	Rickshaw	Apr. 2015 / India
UberMOTO	Motorcycle	Dec. 2016 / Hyderabad
UberBOAT	Boat	Jun. 2015 / Istanbul
UberFreight	Matching truck drivers and freight forwarders	May 2017 / U.S.
...		

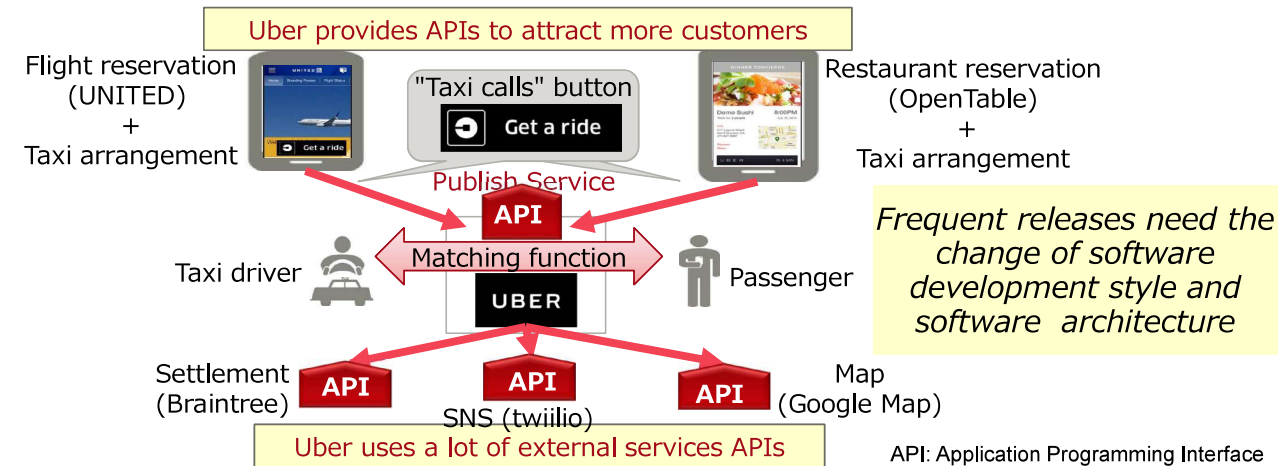
7

Copyright 2020 FUJITSU LABORATORIES LIMITED

## DX Success Case: Frequent Releases of Software



- Efforts to continually improve the value of services
  - Service APIs makes system functions available from the outside and lead to business opportunities.



8

Copyright 2020 FUJITSU LABORATORIES LIMITED



## Software Issues for Achieving DX





9

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Changes in Software Development Style



- Diverse development and operation according to various architectures

Purpose	QCD maximization	Change response	Feedback from operation	For achieving DX
Development Style	Waterfall 	Agile 	DevOps 	DevOps for Microservices 
Features	Sequential phases from requirement definition	Develop flexibly as requirement changes	Feedback from operation to develop	Develop and operate each microservice individually
Integration scope	Only reliable products Use	With the fast-changing OSS (No. of OSS: 28 million)	With OSS and third-party services via APIs (Number of APIs: ProgrammableWeb enrollment 20,000)	
Programming Language	Several (COBOL, Java, C, C #...)	Dozens (Java, C, Swift, C, Scala, JavaScript, Ruby, PHP, Python, Kotlin, Go, *...)		

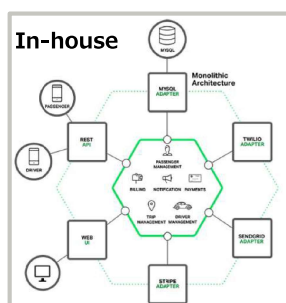
10

Copyright 2020 FUJITSU LABORATORIES LIMITED

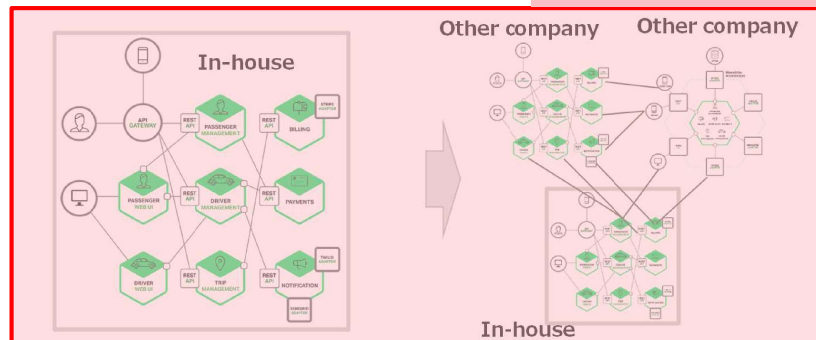
## Changes in Software Architecture



- From monolithic architecture to distribute, loosely coupled microservices architecture
- Complex architectures are difficult to control overall quality **For achieving D**



**Monolithic arch.**  
Operate in large chunks with multiple functions (subsystem)



**Microservice arch.**  
Operate independently for each  
function or organization

**Federated microservice arch.**  
Include services from other companies

11

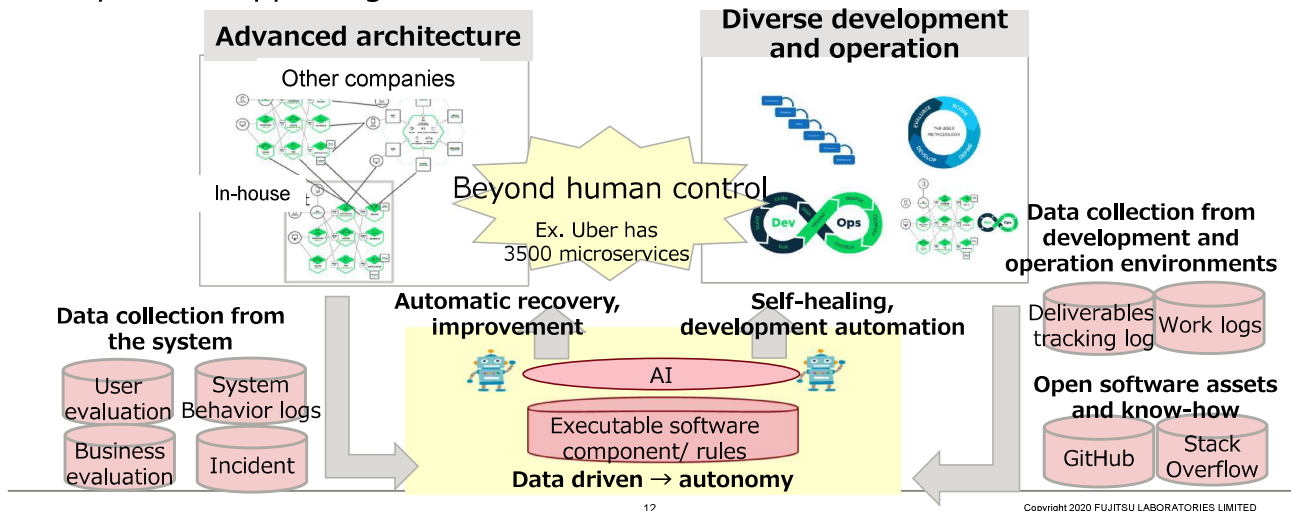
Copyright 2020 FUJITSU LABORATORIES LIMITED



## Overcome advanced architecture and diverse development and operation



- Data-driven software, including AI components, provides “trust” for systems supporting DX

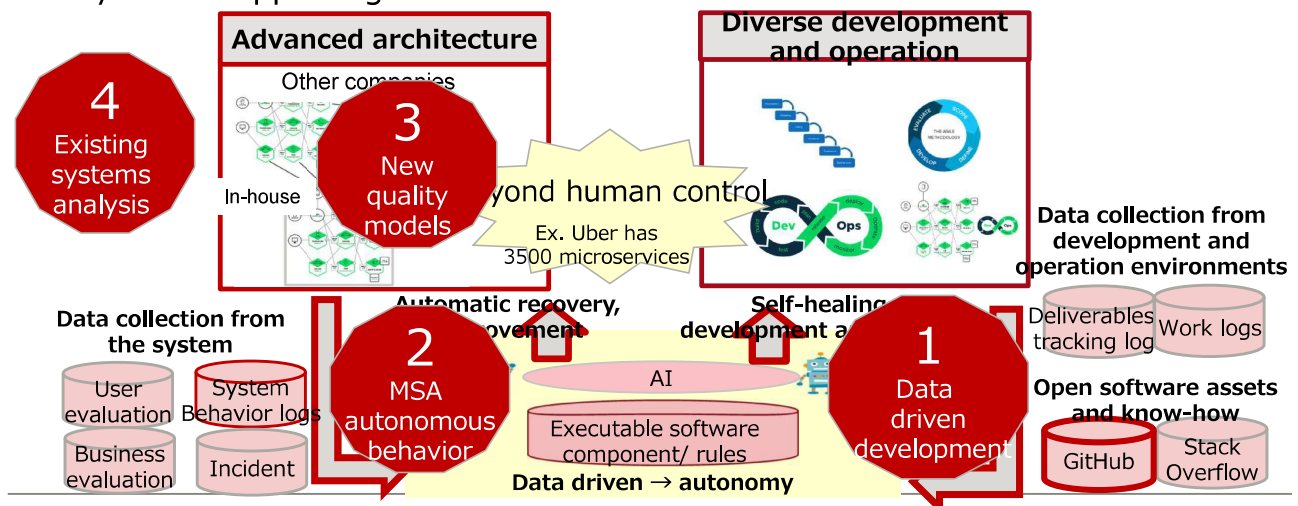


Copyright 2020 FUJITSU LABORATORIES LIMITED

## Overcome advanced architecture and diverse development and operation



- Data-driven software, including AI components, provides “trust” for systems supporting DX



Copyright 2020 FUJITSU LABORATORIES LIMITED

## Key Areas of Trust in DX Systems Development



1. To make good use of development environment data, OSS, and open knowledge  
Ex. Data-driven repair synthesis for static analysis violation
2. Challenge to the autonomous behavior of each microservice  
Ex. MSA(Microservice Architecture) execution infrastructure parameters automatic adjustment
3. New quality models for new software architectures  
Ex. Quality model and quantitative evaluation method for Web API  
Ex. Requirements-driven method to determine quality characteristics and measurements for machine learning software systems (MLS)
4. Technical debt in legacy systems. To promote strategic system renovation while analyzing and evaluating the current state of assets and carrying out classification. (DX report)  
Ex. Existing systems analysis technology

14

Copyright 2020 FUJITSU LABORATORIES LIMITED



## 1. AI Technology to Automatically Recommend Software Repairs

15

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Background

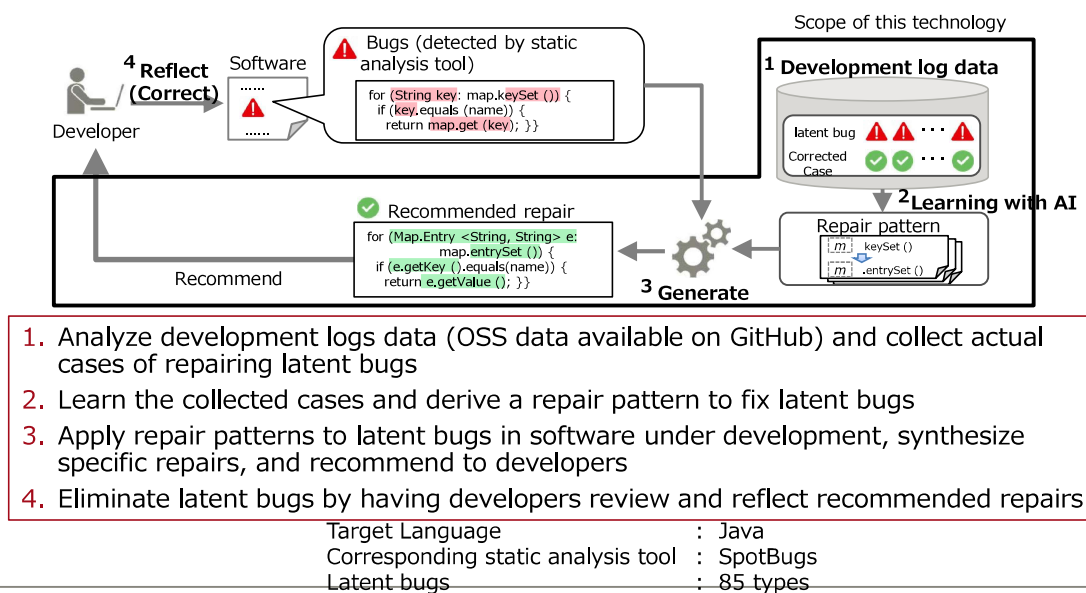


- A technology that leverages AI to automatically generate recommended repairs for latent bugs in software detected by static analysis tools, offering insights to the software developers
- Latent bugs are some parts of the software that may need improvement, including performance degradation and incorrect behavior and poor maintainability
- Hence we suppose the latent bugs can be judged only from the description of the source code without executing the software and the test against it
- It is expected that the use of this technology could shorten software development and maintenance times compared to will develop repair each identified latent bug manually

16

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Technical Overview



17

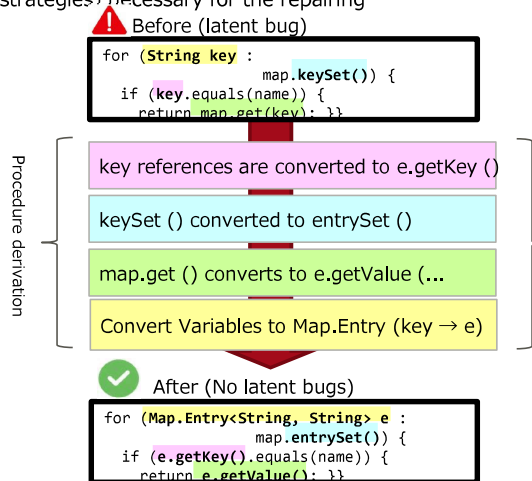
Copyright 2020 FUJITSU LABORATORIES LIMITED

# Learning about Repair Strategies and Synthesis of Repairs



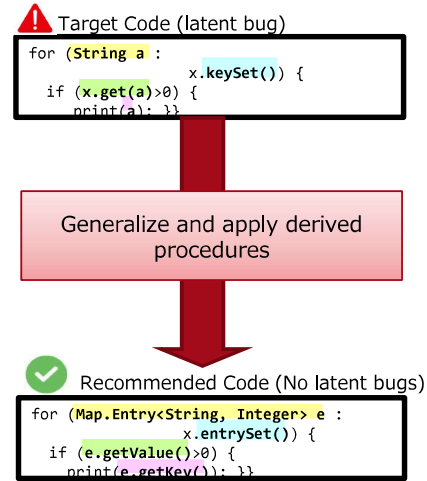
## 1. Learning: Programming by Example

Compare the source code before and after repairing the latent bugs and derive a series of conversion rules (i.e., repair strategies) necessary for the repairing



## 2. Recommendation: Synthesizing of repairs

Generalize the derived conversion rules and apply them to the target code to synthesize the repairs



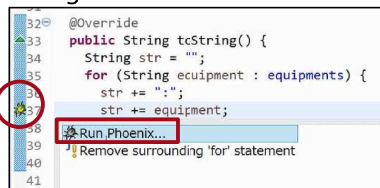
18

Copyright 2020 FUJITSU LABORATORIES LIMITED

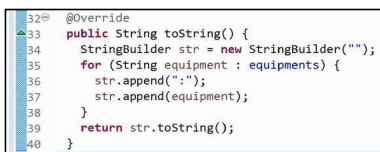
## Demo (Screenshot)



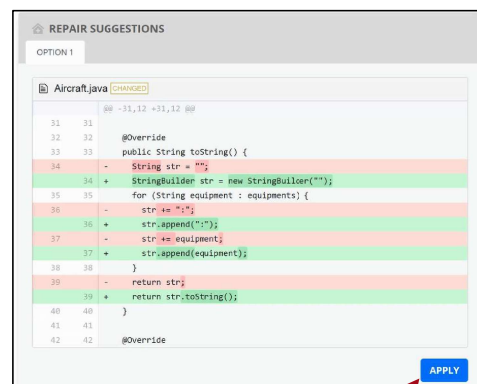
1. Run this tool on the line with a latent bug



3. Once approved by the user, apply the modifications to the source code being edited



2. Recommend source code modifications



Apply button

19

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Demonstration



- Demonstration: A flight reservation site
  
- Evaluation with our customer
  - The system that handles financial transactions
  - This technology recommends appropriate repairs for over half of the latent bugs detected
  - Using these proposed repairs, it would have been possible to reduce the time required to repair the latent bugs by up to 30% compared with performing the task manually



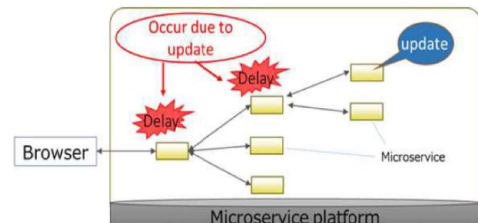
## 2. A Technology for Automatically Adjusting the Configuration Parameters of Microservices

## Background



- Systems that consist of multiple microservices can be independently developed and operated to meet demands according to service requirements

- Problem: Changes to a single microservice could delay or otherwise impact the entire system.  
To ensure service quality, developers had to validate the entire system with every update.



- Goal: By automatically adjusting parameters to guarantee proper operation of the entire system, users can easily maintain service quality even when making changes to microservices

22

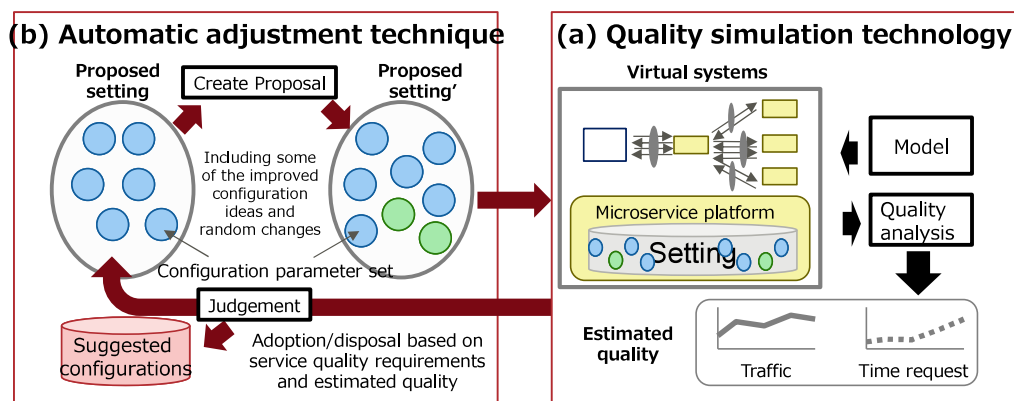
Copyright 2020 FUJITSU LABORATORIES LIMITED

## Developed Technologies (1/2)



### a. Simulation technology to estimate service quality

- Based on queueing theory, the technology uses mathematical models to analyze waiting lists and mitigation measures to model communication between microservices and simulate the behavior of the system overall



23

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Developed Technologies (2/2)



- b. Automatic adjustment technique to search for optimal parameters
  - By combining genetic algorithm and quality simulation techniques
  - A Technology that searches for the optimal combination of configuration parameters, such as CPU/memory allocation, and number of retries, given the quality requirements of the services as constraints
- Outcome
  - An in-house streaming service built with microservice platform Kubernetes and Istio has been used to confirm that parameter adjustments configured with the newly-developed technology can reduce the time required from 3days for manual revision to just 1hour



## 3. New Quality Models and Characteristics

## 3-1 Quality Model and Quantitative Evaluation Method for Web APIs



- Background:
  - Web APIs are a core asset in the development of enterprise information systems
  - Web APIs have following features different from the conventional system APIs
    - ① RESTful Interface Definition
    - ② Evolution at Runtime
  - We need to discuss the impact of the quality imposed by these features

26

Copyright 2020 FUJITSU LABORATORIES LIMITED

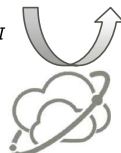
### Problem Context (1)



[Restful Interface Definition]

- Representation of resources that conform to the principles of REST
- Independent of programming language
  - ⇒ no type checking, defined manually

Get /Request URI



```
{
  "products": [
    {
      "upfront_fare_enabled": true,
      "capacity": 4,
      "product_id": "d4abaae7-f4d6-4152-91cc-775",
      "price_details": {
        "service_fees": [

```

- Conventional system APIs: using a programming language
  - ⇒ type checking mechanism,
  - tools for generating an interface definition, ...
- We define new “**Learnability**” to solve this problem
- Motivation: To estimate unconsidered cost for learning the usage of Web API

27

Copyright 2020 FUJITSU LABORATORIES LIMITED

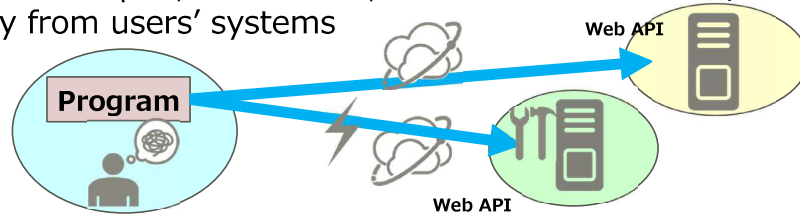


## Problem Context (2)



[Evolution at Runtime]

- Web APIs are developed, maintained, and executed remotely and independently from users' systems



- Conventional system APIs: can control the change of the APIs since the system APIs run on the users' systems.
- We define new "**Interoperability**" to solve this problem
- Motivation: To estimate the maintenance costs imposed by the Web APIs in advance

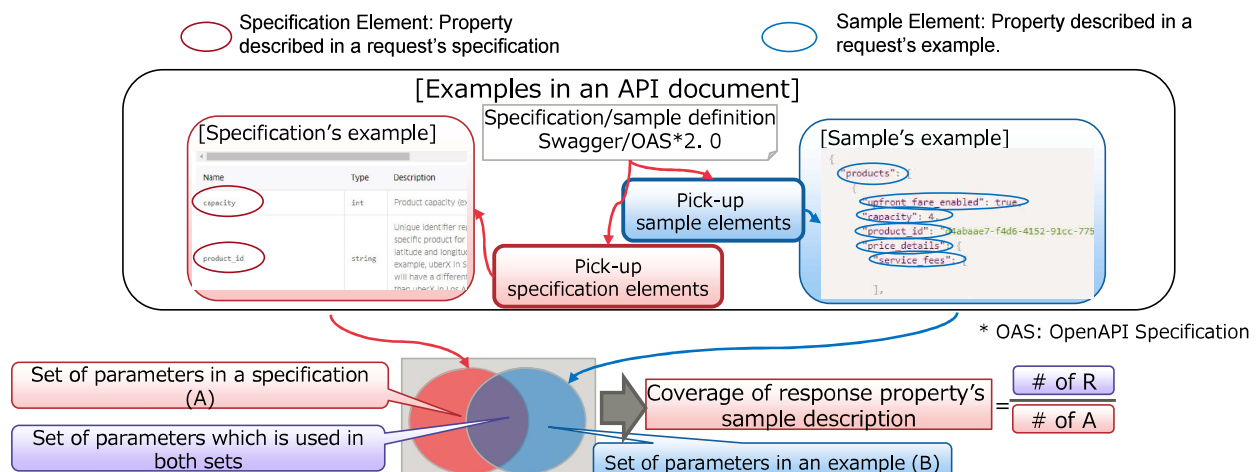
28

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Example of Learnability: How to Calculate "Coverage of response property's sample description"



- The ratio of the number of sample elements for the number of specification elements



29

Copyright 2020 FUJITSU LABORATORIES LIMITED

# Quality Model: Interoperability

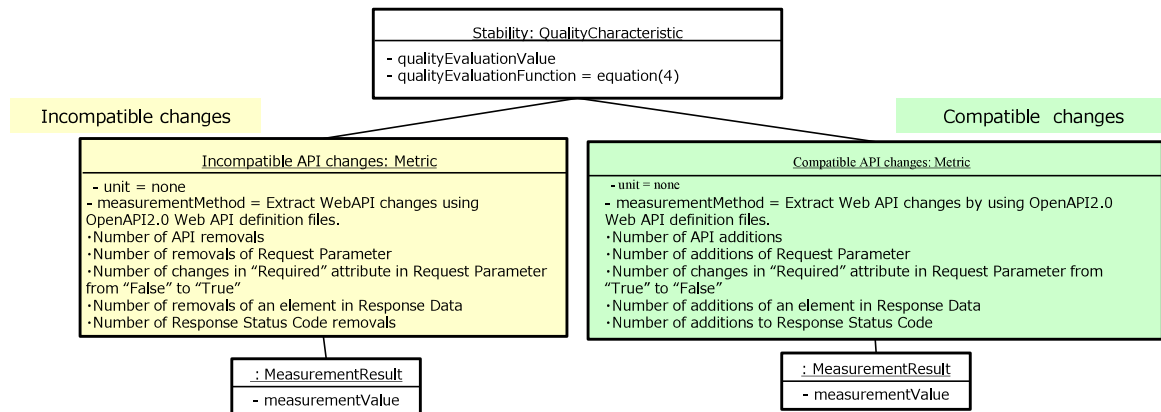
$$\text{Web API Interoperability} = \frac{1}{1 + \sum_{s=1}^{|S|-1} (k_i I(s) + k_c C(s)) h w(s)}$$

Temporal impact

Incompatible changes

Compatible changes

Collection of snapshots



30

Copyright 2020 FUJITSU LABORATORIES LIMITED

## 3-2 Quality Characteristics and Measurements for Machine Learning Software

### ■ Background

- Machine learning software systems (MLS): A software system incorporating machine learning algorithms.
- Unique Quality of MLS: MLS has unique quality based on the nature of machine learning. It is different from the quality of conventional software systems.

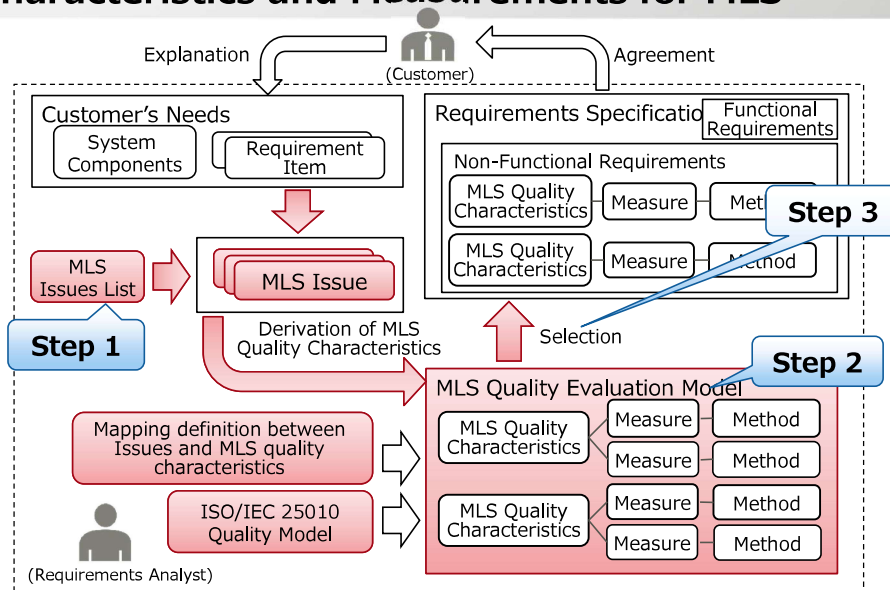
### ■ Problems in MLS development

- Difficult to predict the actual behavior of the learning model.
- Agreement with customers is often made through a PoC (Proof of Concept) to confirm the behavior in advance.
- The quality of the learning model is considered in PoC, but quality of MLS is not enough considered.
- Quality model and characteristics of MLS are neither established nor standardized

31

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Approach: Requirements-Driven Method to Determine Quality Characteristics and Measurements for MLS



32

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Proposed Method



- Proposed method consists of 3 steps

**Step 1: Select MLS Issues and define MLS issue List on quality**

MLS issue List      MLS Issue      MLS Issue

**Step 2: Define MLS quality evaluation model (MLS quality characteristics and measures)**

MLS Quality Evaluation Model      MLS Quality Characteristics      Measure      Method

**Step 3: Selection method of MLS quality characteristic and measures based on issues**

33

Copyright 2020 FUJITSU LABORATORIES LIMITED

## 4. Existing Systems Analysis Technology

34

Copyright 2020 FUJITSU LABORATORIES LIMITED

### Problems of Existing Systems

Point of view	Problem area	Examples of specific problems
Technical obsolescence	<ul style="list-style-type: none"> <li>Infrastructure hardware</li> <li>OS Platform</li> <li>Program language and framework</li> </ul>	<ul style="list-style-type: none"> <li>Decrease in the number of engineers</li> <li>Vendor lock-in</li> <li>High maintenance and operation costs</li> <li>Difficulty in responding to new technologies and linking with other systems</li> <li>Out of Support</li> </ul>
Enlargement and complexity	Application program	<ul style="list-style-type: none"> <li>Complex dependencies between functions</li> <li>Growth due to increased size and copy development</li> <li>Increase in non-operating assets</li> <li>More inactive features</li> <li>Slow Release</li> </ul>
Black is boxing	Application program	<ul style="list-style-type: none"> <li>Personalization of maintenance</li> <li>Loss of knowledge due to transfer of personnel in charge</li> <li>Narrowing down problem areas</li> <li>Absence of people capable of defining requirements</li> </ul>

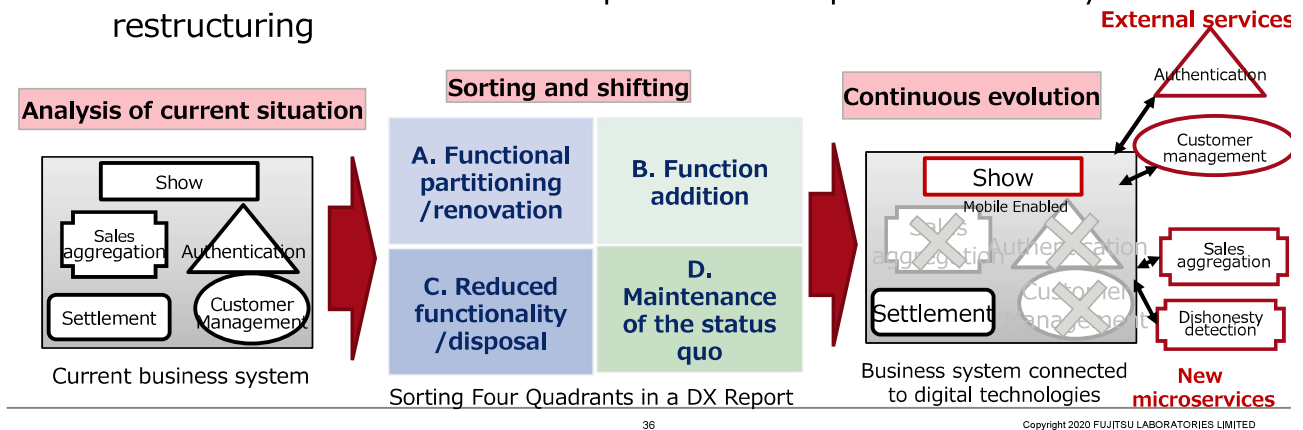
35

Copyright 2020 FUJITSU LABORATORIES LIMITED

## DX Execution Process with Existing Systems



- For DX, it is necessary to promote strategic system renovation while analyzing and evaluating the current state of assets and carrying out sorting (DX Report)
- Evaluate each function in four quadrants and plan for future system restructuring

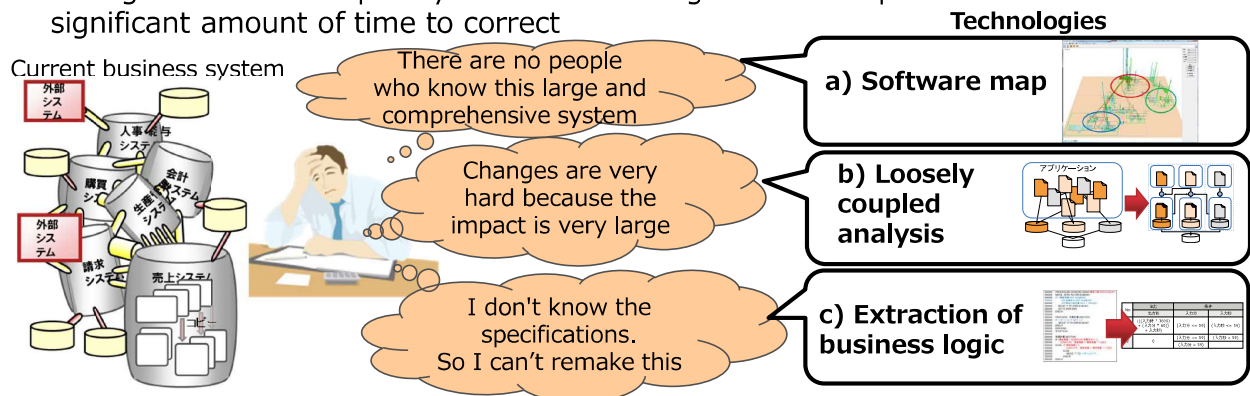


36

## Typical Existing System Situation



- "Black box" : The specifications have not been updated due to many years of renovation, and maintenance work depend on people.
- Enlargement and complexity : Effects of changes are widespread and take a significant amount of time to correct



Technologies to assess current situation and assist sorting

37

Copyright 2020 FUJITSU LABORATORIES LIMITED

## a. Software Map



### ■ Summary

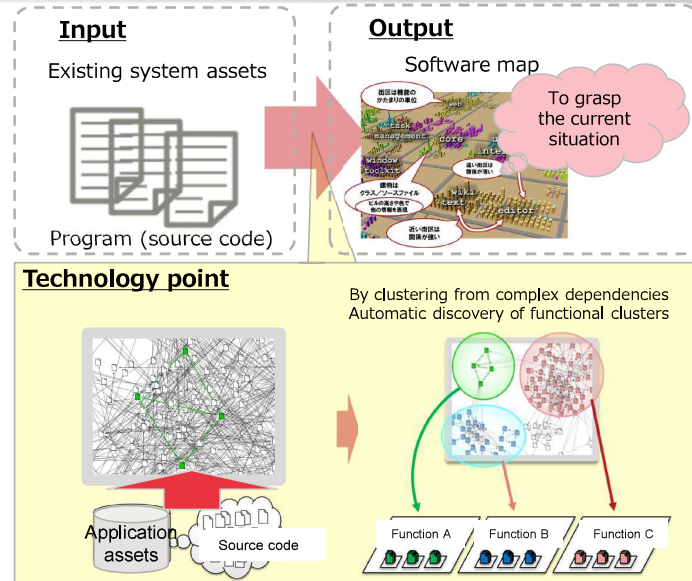
- Analyze complex, large-scale business software to create a map of the structure.
- You can objectively evaluate the current situation without knowing the program

### ■ Feature

- Program dependencies are scored and deeply related parts are automatically grouped
- Display the extracted groups in a map format and visualize the whole with a high degree of abstraction

### ■ Apply Scene/Effect

- Clarify the gap between the current state and the structure at the time of design
- Identification of maintenance problems
- Consideration of transition policy



38

Copyright 2020 FUJITSU LABORATORIES LIMITED

## b. Loosely Coupled Analysis



### ■ Summary

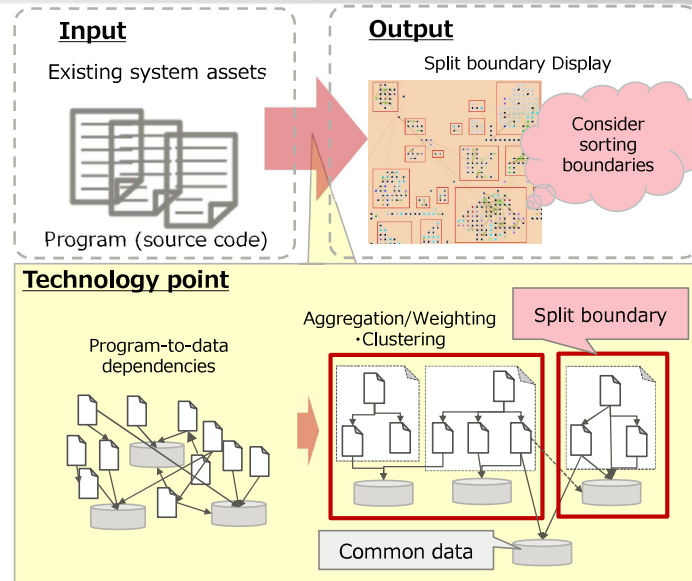
- Analyze complicated business systems and automatically present boundaries that are easy to divide to support functional division

### ■ Feature

- Static and dynamic analysis of program and data dependencies to extract boundaries that minimize interdependencies
- Visualize using software map technology to locate areas with strong relationships close to each other

### ■ Apply Scene/Effect

- Improved maintainability of existing systems
- Stepwise migration of large systems
- Cutting out functions and making them as web services



39

Copyright 2020 FUJITSU LABORATORIES LIMITED

### 3. Business Logic Extraction technology

#### ■ Summary

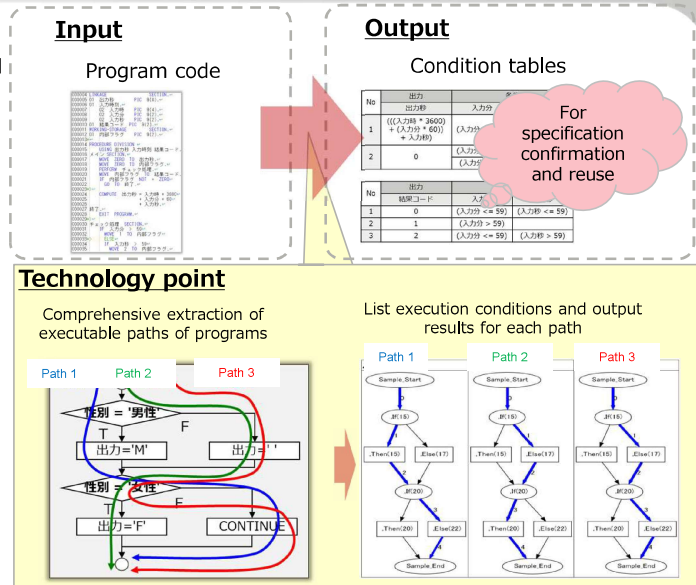
- Extracting the business specifications implemented in the program as condition tables those are easy to understand and support the understanding and reuse of specifications.

#### ■ Feature

- "Symbolic execution" (the value of the variable in the program as a symbolic value) performs all program execution paths and execution conditions
- By dividing or narrowing the analysis target, we avoid explosion of execution paths and accommodates large programs

#### ■ Apply Scene/Effect

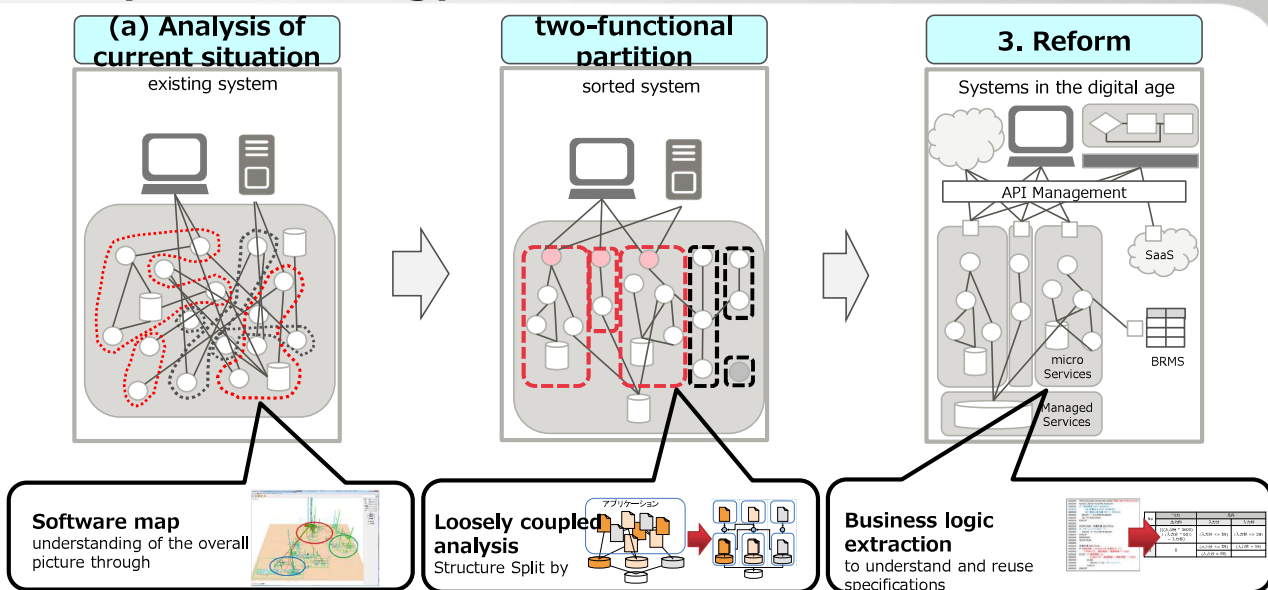
- Rearrangement of specifications of the current system
- This function describes the business logic in the form of rules. Then customers can move them to BRMS, which can execute them without programming.



40

Copyright 2020 FUJITSU LABORATORIES LIMITED

### Image of Migration using Existing Systems Analysis Technology



41

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Improving Experience Technologies - An Example of New Services with DX -

42

Copyright 2020 FUJITSU LABORATORIES LIMITED

### Concept Video



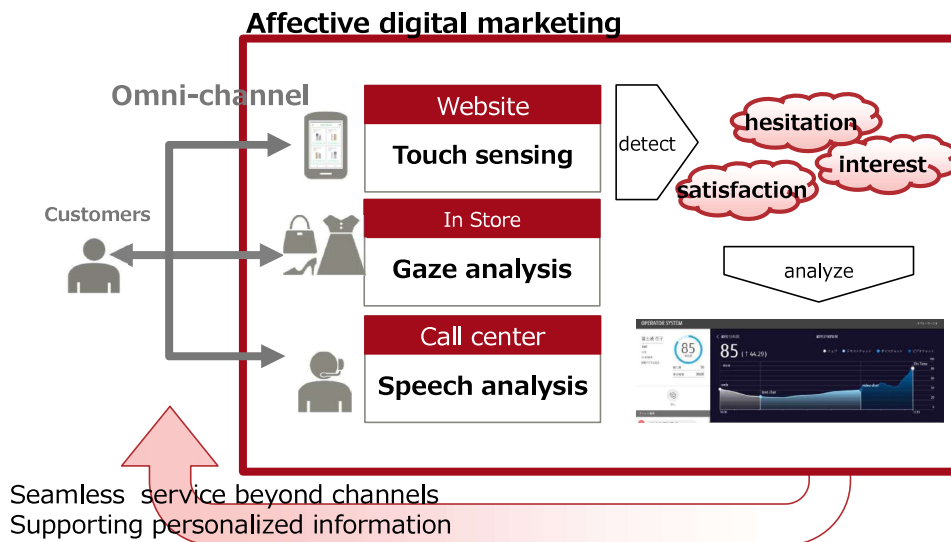
43

Copyright 2020 FUJITSU LABORATORIES LIMITED



# Overview of concept

FUJITSU



44

Copyright 2020 FUJITSU LABORATORIES LIMITED

## Technologies

FUJITSU

- Estimation of **customer's experience(CX)** such as interest, hesitation and satisfaction for digital marketing.

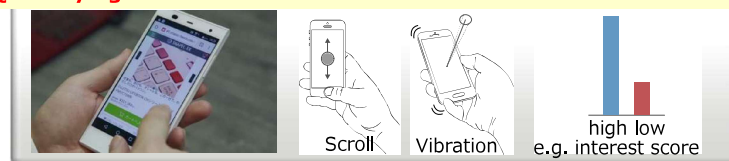
Quantifying customer satisfaction in conversation



Quantifying customer interest and hesitation in stores



Quantifying customer interest and hesitation at ecommerce website



45

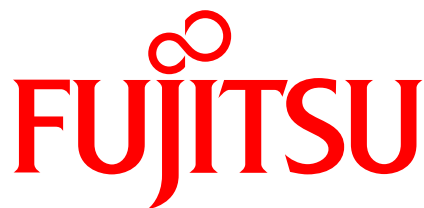
Copyright 2020 FUJITSU LABORATORIES LIMITED

## Summary



- DX(Digital Transformation) and its Success Cases
  - Frequent releases of software
- Frequent releases need the change of software development style and software architecture
  - To overcome advanced architecture and diverse development and operation, data-driven software, including AI components, provides “trust” for systems supporting DX
  - Key Areas of Trust in DX Systems Development: Data driven development, MSA autonomous behaviors, New quality models, Technical debt in legacy systems
  - New issues: Improving experience technologies
- Software technologies are important for achieving DX

Thank you for your attention !



shaping tomorrow with you

Session 2:  
Sensing and Analysis  
( Chair: Yoshitaka Nakamura )



# To Estimate a Specific Position Related to an Event

Takuma Toyoshima<sup>\*</sup>, Takuo Kikuchi<sup>\*\*</sup>, Masaki Endo<sup>\*\*</sup>, Shigeyoshi Ohno<sup>\*\*</sup> and Hiroshi Ishikawa<sup>\*\*\*</sup>

<sup>\*</sup>ICT Production Support, Polytechnic Center Shiga, Japan

<sup>\*\*</sup>Division of Core Manufacturing, Polytechnic University, Japan

<sup>\*\*\*</sup>Graduate School of System Design, Tokyo Metropolitan University, Japan  
{kikuchi, endou, ohno}@uitech.ac.jp

**Abstract** - Sensors can be used to track the movement of a person at an event. Because physical sensors are expensive to install and maintain, users on SNS can be treated as sensors instead of physical sensors to observe real-world events. Therefore, many data are required. However, from the viewpoint of information protection, few SNS have accurate location information. Therefore, we assess a method of estimating position information related to a specific event. The estimation accuracy was evaluated using actual data tweeted in Chiyoda ward, Tokyo.

**Keywords:** Deflation structure, Location estimation, Twitter, Real time analysis, Social sensor

## 1 INTRODUCTION

In modern society, with computerized devices and systems, many means exist to acquire information of various types in real time. One means is Twitter<sup>1</sup>, a microblogging service that shares short sentences called "Tweets" of 140 or fewer characters. It is widely used throughout the world, as it is in Japan. Many users regard it as a medium by which they can post recent information casually. Posting of location information can be done easily with so-called geotags via a smartphone. Therefore, it is a social medium that can immediately notify many people of what is happening and where. Based on these characteristics, such a social medium is anticipated for use as a social sensor for observing the real world without using expensive physical sensors [1].

Social sensors can reveal a situation in real time even if one is not present on the scene. For example, if one can guess the best time for cherry blossoms and autumn leaves before visiting a place, one could actually go there without any concern that cherry blossoms have not yet bloomed or that they have already fallen. If public transportation is halted and one knows that people are crowding into stations, then one can avoid those crowds. If a person knows that congestion in an area has eased, then the person could stop avoiding the area. If a reveler wants to attend a Halloween party in Shibuya, then that person would want to hurry while the party is still exciting. After the Halloween party has settled down, it would not be so interesting. Alternatively, to avoid a raucous party, one might wait until after it has

settled. Social sensors would be useful to inform people to make such choices.

Analyzing today's events using yesterday's data is not always helpful, but predicting the movements of other people in a specific place in real time can help a person decide whether to visit a certain place or not. Assuming that one is not actually present in a certain place, Twitter data can be useful to ascertain the concentration of people in real time while avoiding deployment of expensive sensors. This study was conducted to produce a means of real-time estimation of human motion by analyzing geotagged tweets.

## 2 RELATED WORK

Kleinberg [2] proposes a method for modeling text stream bursts and for extracting structures. This method is based on modeling a stream using an infinite state automaton. A salient benefit of Kleinberg's approach is that it can represent burst duration, degree, and weight for each topic. Therefore, it is used widely for various applications. Nevertheless, it is unsuitable for real-time burst detection because analyses cannot be done immediately for occurrence of a certain event.

Studies conducted by Y. Zhu and D. Shasha [3] [4] and by X. Zhang and D. Shasha [5] examine bursting algorithms that monitor bursts efficiently over multiple window sizes. These techniques enable near real-time burst detection by shortening of the monitoring interval. However, they require monitoring of the number of occurrences of events at regular intervals. Data must be stored even if no event has occurred.

Ebina et al. propose a method for real-time burst detection [6] [7], which is achieved by inference of whether or not each event (each tweet posting) is a burst. The number of calculations is reduced by compressing data held at the time of occurrence of concentrated events. Burst detection with high real-time capability is achieved, it remains unclear whether the burst state continues or immediately ends solely by the burst occurrence.

Endo et al. use the moving average to make full-fledged decisions [8] [9]. This method detects burst occurrence and burst state continuation and convergence. However, because the tweet occurrence frequency is used with a fixed window size, real-time properties are quantized by the window size. Using the Tweet Posting frequency requires setting of a

<sup>1</sup> Twitter, URL<<https://Twitter.com/>>

certain time interval for frequency calculation. This fixed interval impairs real-time performance.

Large amounts of tweet data are necessary to estimate people's movements in real time. However, few tweets include *any* location information. Furthermore, much location information is ambiguous. Moreover, tweets with *accurate* location information are even fewer. Therefore, research is being conducted to obtain location information from tweet contents [10] [11]. Methods have been designed to obtain location information by analyzing the vocabulary included in the tweet text. Such methods identify a target, such as an event or building, in a tweet that has no location information in the first place. Therefore, it is inferred that the tweet was tweeted from the event venue or from the position of the object. However, for the present study, we strive to identify and use more accurate position information estimation. Even if a reference to an event exists, tweets made before going to the event venue or after returning home are not tweets that were issued from the event venue. Nevertheless, removing such tweets from the overall data is not easy.

### 3 TARGET EVENT AND DATA

The target event for this experiment was the visit of the General Public to the Imperial Palace after the Accession to the Throne on May 4, 2019. About 141,000 people visited as members of the general public. Their Majesties the Emperor and Empress appeared at the balcony of the Chowa-Den Hall six times to greet visitors who had gathered there. Participants were able to enter from the main gate of the Imperial Palace. The time from 9:30 am to 2:30 pm was the entry time. Analyzed tweets with geotags were made in the area around the Imperial Palace. Those in this range were visitors of the general public who tweeted. They tweeted while they were waiting or after they left. We can imagine that they would be unable to tweet when moving to Chowa-Den immediately before each appearance, and that they would refrain from tweeting during each appearance. By checking the tweet status, one can estimate the participants' movements: whether waiting or moving.

The tweets to be analyzed were tweets including geotag "coordinates". The geotag "coordinates" data represent the location at which the terminal used for posting a tweet is represented by single latitude and longitude coordinates. Therefore, is the data are highly useful as positioning data. Tweets were extracted during 00:00:00 – 23:59:59 on May 4, 2019. The tweet extraction range was found empirically for the following four latitude and longitude ranges.

[35.677002, 139.753658]

[35.689604, 139.753658]

[35.689604, 139.761212]

[35.677002, 139.761212]

By filtering using the conditions shown above, 198 tweets were extracted. Of these, 116 accounts were tweeted.



Figure 1: Target Area(in front of the Imperial Palace).

Table 1: The correct answer to estimate.

	time	action of people	
before open gate	– 9:40	Stay	May 4
around appearance	9:40 – 10:10	Move	Deflation
	10:10 – 10:40	Stay	
	:	:	
	14:10 – 14:40	Stay	
around appearance	14:40 – 15:10	Move	Deflation
	15:10 – 16:00	Stay	
event end	16:00 –	Converge	Deflation

From the appearance time, we inferred the time period during which people were thought to have moved (or stayed), as shown in Table 1.

### 4 DETERMINATION METHOD

The following two methods are used as the deflation determination method. One is a method based on the method reported by Endo et al. They succeeded in determining changes in people's posting on a daily basis, such as cherry blossom viewing time estimation. In our study, the movement of a person is judged in units of minutes instead of days. The other is a method derived from the real-time burst determination method reported by Ebina et al. Our criteria are reversed to determine a deflation rather than a burst.

#### 4.1 Method Based on the Endo et al. Method

This method uses the tweet posting frequency in a reciprocal relation with the Tweet posting interval. Tweet post intervals have high real-time characteristics because they depend on each tweet. However, a certain time interval

must be found to calculate the tweet posting frequency. Usually, the real-time performance is low because the time interval is larger than the tweet posting interval.

The Endo et al. method uses a moving average of the frequency of posting tweets to estimate the full bloom of cherry blossoms or other phenomena. The method calculates the frequency daily and examines differences between the 5-day moving average and the 7-day moving average.

The judgment criterion for the best time is when the 5-day moving average becomes greater than the 7-day moving average and becomes larger than the average of the prior year. A comparative experiment was conducted using this condition. However, the tweet posting frequency is not calculated on a daily basis, but on a 5-min basis. Unlike efforts to infer the best time to see cherry blossoms and so on, we wanted to assess the movement and congestion of crowds of people. Therefore, a short time interval was used.

From the change in the moving average of the tweet posting frequency, it is examined how accurately the movement of the person in Table 1 can be determined. Table 2 shows results of the quantitative evaluation. Not only judgment based on the difference between 5 moving average and 7 moving average, but also judgment result when using the difference between 3 moving average and 5 moving average and judgment result when using the difference between 3 moving average and 10 moving average also evaluated. The precision is high, but the recall is low. Because moving averages are used, it is not possible to respond sensitively to changes; also, there are many oversights. Both of those shortcomings lead to poor recall.

Table 2: Evaluation using the Endo et al. method.

	Precision	Recall	F-value
3(15-minute) moving average / 5(25-minute) moving average	47.06%	5.76%	10.26%
5(25-minute) moving average / 7(35-minute) moving average	64.81%	25.18%	36.27%
3(15-minute) moving average / 10(50-minute) moving average	48.94%	16.55%	24.73%

## 4.2 Method Based on the Ebina et al. Method

The Ebina et al. method uses the tweet posting interval instead of the tweet posting frequency for real-time determination. Similarly to assessment of the change of the moving average, burst judgment is performed by the change of multiple tweet posting intervals. Using this method, deflation is inferred by reversing the judgment conditions. The deflation occurrence condition is when the tweet posting interval changes longer than before.

Similar to the previous section, we examined how accurately the deflation of the person in Table 1 can be judged under the above deflation judgment condition. Table

3 shows quantitative evaluation results. The Ebina et al. method recall rate is high because it reacts in real time. Its precision is not as good as that achieved when using the Endo et al. method, but the recall and F-value are high.

Table 3: Quantitative evaluation according to Ebina et al.

	Precision	Recall	F-value
5 number analysis	36.46%	55.56%	44.03%
10 number analysis	42.98%	77.78%	55.37%
15 number analysis	39.51%	50.79%	44.44%

## 5 LOCATION ESTIMATION

We intend to improve the accuracy of the proposed method by increasing the dataset used for deflation determination. We estimated people's movements using tweets with geotag "coordinates" that can obtain accurate location information. However, only 116 accounts posted the tweets used in the experiment. The number of visitors conducting a general visit is 141,130. Even if the percentage of users who tweet is low, one can infer that tweets are actually posted from more accounts because the data are limited to those with geotag "coordinates" that can specify the position. Therefore, we perform verification by increasing the number of analysis targets using tweets with unclear positioning. Specifically, the position is estimated using machine learning from a group of tweets with "place" that represents Chiyoda ward. The number of tweets to be analyzed is therefore increased. Then the accuracy of deflation determination is evaluated quantitatively. First, we discuss extraction of tweets posted during the general visit from tweets that are clearly posted in Chiyoda ward.

Compared to the number of tweets with geotag "coordinates", the number of tweets including only the geotag "place", which records rectangular range information represented by four latitudes and longitudes, is about four times as large. The tweets that are useful as analysis targets are estimated from the group of tweets to which only the geotag "place" was added. Those tweets have ambiguous position information, but they are added as analysis targets. Similarly to the case of analysis using only tweets with geotags "coordinates", recall, precision and F-value are calculated. Then their accuracy is evaluated quantitatively.

### 5.1 "Place" Data

As a tweet with a geotag for which position information is ambiguous, a tweet for which the geotag metadata is "place\_type:city" is used. The data to be classified by machine learning are tweets with no geotag "coordinates" added on May 4, 2019. Only "place" data representing Chiyoda ward, Tokyo are used. The Imperial Palace is located there. There were 3132 tweets.



The attached “place” data were confirmed for the 198 tweets with “coordinates” used in the previous chapter. Results show that 198 “place” data were all the same value. The following four points were recorded.

[35.6686, 139.73]  
[35.7052, 139.73]  
[35.7052, 139.783]  
[35.6686, 139.783]

Figure 2 portrays these four points. The range including this place is the range surrounding the entire Chiyoda ward.



Figure 2: “place” of Chiyoda Ward.

Tweets that have only the geotag “place” are shown in the range of latitudes and longitudes of the four points. They can be narrowed down to the municipality, but the tweeted position cannot be found. Therefore, we analyze the tweet contents using natural language processing and consider a method to estimate the user's position more accurately based on the tweet contents. To infer the location, 3132 tweets with the geotag “place” representing Chiyoda ward, where the Imperial Palace is located, are binary-classified using machine learning to estimate whether or not the tweet is from a user who has visited to the Imperial Palace. When classifying tweets with only geotag “place” added by machine learning, teacher data are extracted from text data of tweets with geotag “coordinates” added. In the classification, the target data for estimating the position were text data of the tweet to which only the geotag “place” representing Chiyoda ward was added.

## 5.2 Vectorization and dimension reduction

For using machine learning, we create teaching data consisting of a set of tweets as a model. The model includes text data of tweets with geotag “coordinates”. All text data of model tweets are collected to compile a word dictionary. The word dictionary comprises noun words obtained from all words that appear by analyzing the tweet set morphologically and dividing it into words. This time, we extracted only nouns. Then, considering that the absolute number of tweets with the geotag “coordinates” is small, we

particularly examine nouns for each tweet. MeCab is used to extract the morphemes. Furthermore, by normalization, character strings including only numbers, katakana, and alphabet characters are excluded as stop words. For dictionary data used in MeCab, in addition to the IPA dictionary provided as standard in the morphological analyzer, a user dictionary created from keyword files of “Wikipedia” and “Hatena Keyword” is also used so that minor nouns can be supported.

To convert text data into numerical data that can be processed using machine learning, the data must be vectorized. Bag of Words (BoW) is used for vectorization. With BoW, the number of times a word appears is counted for each tweet. A matrix is generated from the counted number of words.

There is a great amount of noise. The machine learning result is affected by vectorizing the words extracted using morphological analysis. Two preprocessings, TF-IDF and LSI, were done to use the feature vector extracted from each tweet content as optimum data for use in machine learning.

The vectorized features are weighted by TF-IDF, which is a method to weight the words when classifying individual tweets when the number of occurrences of highly important words is high in a tweet set. Term Frequency (TF) represents the number of times a word appears in a tweet. Document Frequency (DF) represents the number of tweets in which a word appears. Also, IDF is the logarithm of the reciprocal of DF.

Latent Semantic Indexing (LSI) is a dimensional compression method using Singular Value Decomposition (SVD). The LSI method specifically examines the latent meaning of words for mitigating over-learning and for reducing learning costs. Indexing synonyms and making synonyms into a vector can be done by indexing the latent meanings of words.

Regarding the number of dimensions to be compressed, the greater the number of dimensions used for machine learning becomes, the higher the calculation cost becomes. Moreover, the processing time increases. As described in this paper, we reduce the dimensions of feature vectors weighted by TF-IDF to 100 dimensions by LSI.

## 5.3 Incorrect answer area

To estimate the place from which a tweet with only a geotag “place” was posted, we used SVM, which is a learning model capable of binary classification by machine learning. The place is classified by SVM from the tweet contents. Actually, SVM has good compatibility with binary classification and high generalization performance. The following ranges for extracting correct and incorrect data are both included in the “place” range (Fig. 2) for Chiyoda ward.

Among the teaching data, the 198 tweets used in the experiment in the previous chapter were used as correct answers. The tweet group in the range where the event occurred is used as correct answer data.

Incorrect answer data was extracted from tweets in areas where the event did not occur. These tweets have a “coordinated” geotag. The range of the incorrect answer data is the same area as that of the correct answer data. This is



the location of the target event, 200 meters east of the Imperial Palace. The range of latitude and longitude from the southwest is given below. Figure 3 portrays the range of the extracted incorrect data on the map.

[35.677002, 139.763425],  
[35.689604, 139.763425],  
[35.689604, 139.770979],  
[35.677002, 139.770979]



Figure 3: Area of incorrect data.

By extracting the tweets to which "coordinates" were added under the conditions depicted in Fig. 3, 866 tweets were eventually extracted. Then the text set that combines correct and incorrect data above is used as a learning model.

## 5.4 SVM

We used a classification method by Support Vector Machine (SVM) capable of binary classification position of a person who was at the scene of a specific event or trouble from the post to which only the geotag "place" was added. Actually, SVM is adopted as a method to estimate the user attributes and position from tweet contents in many previous studies. For the SVM kernel, we adopt a linear kernel that is often used when classifying large-scale data, sparse data, and text data. In addition, cost parameter C is set to the default value of 1 for learning.

The linear kernel SVM is a classifier that constructs a hyperplane that maximizes the shortest distance (margin) between the classification boundary and the training data.

For machine learning model evaluation, cross validation is undertaken by dividing the data into an arbitrary number K using the Stratified K-Fold method. The tweet structure of the evaluated model is the following. It consists of data of two types. One is 198 tweets with geotag "coordinates" extracted in the correct answer range (Fig. 1), which is regarded as having caused people to stagnate and flow because of the occurrence of events. The other is 866 tweets with geotag "coordinates" extracted in the range (Fig. 2) where the target event (Visit of the General Public to the Imperial Palace after the Accession to the Throne) has not

occurred. The 1064 tweet data, which include these two types of data, are divided into five portions: 4/5 are training data; 1/5 are test data. The tweet vector generated from the training data is the explanatory variable. However, with classification by SVM, for the objective variable, a binary label assigned to the test data is 1 for a correct answer and -1 for an incorrect answer. The training data and test data are exchanged. Classification is performed using SVM five times in total. The average value of the results of five cross-validations was used for estimating the model accuracy. The machine learning library scikit-learn was used to implement SVM. This time, K=5 split cross validation was performed. The average value of the classification correct answer rate for five times was calculated. The classification accuracy rate is an index showing how well the classifier can classify. The classification accuracy rate can be expressed as equation (1).

$$\text{Classification correct answer rate} = \frac{\text{Number of successful classifications}}{\text{Number of evaluation cases}} \quad (1)$$

As a result, it was 80.64%.

## 5.5 Estimated result

The classification target is 3132 tweets with only the geotag "place" that represents Chiyoda ward. The SVM has assigned a label of 1 to tweets that are judged to be in the correct answer range, and a label of -1 to tweets that are judged to be in the incorrect answer range. From extraction of the tweets with correct labels, 1750 tweets were output as correct answers.

These 1750 tweets are combined with data of 198 correct tweets with the geotag "coordinates". Using 1948 tweets, we conduct analysis using the Endo et al. method and the Ebina et al. method. By calculating the precision, recall, and F-value, the accuracy of deflation judgment can be evaluated quantitatively when analysis is performed by adding tweets estimated as having been posted in the range presented in Fig. 1 by machine learning.

Similar to Table 2, Table 4 shows the results of quantitative evaluation using the method of Endo et al. As in Table 2, the three types of parameters are shown. In Table 2, we evaluated only 198 tweets with coordinate data, but in Table 4, as a result of position estimation, the number of tweets used for evaluation is 1948, which is larger than in Table 2. The number of tweets has increased, but the precision and recall have not improved. Table 5 shows the results of the quantitative evaluation using the method of Ebina et al. In Table 5, the number of tweets used for evaluation is larger than in Table 3. Compared to Table 3, the recall is not very good, but the precision is good.

We were able to increase the number of tweets used to judge deflation by machine learning from 198 to 1948, nearly 10 times. The tweet posting position was estimated using SVM, but its classification accuracy rate in cross-validation is about 80%, and it contains about 20% noise. Even if the number of tweets used to judge deflation is increased by about 10 times, noise is included. Therefore, the accuracy of deflation judgment may or may not improve.

Table 4: Evaluation using the Endo et al. method.

	Precision	Recall	F-value
3(15-minute) moving average / 5(25-minute) moving average	48.89%	15.83%	23.91%
3(15-minute) moving average / 10(50-minute) moving average	43.75%	25.18%	31.96%
5(25-minute) moving average / 7(35-minute) moving average	47.37%	19.42%	27.55%

Table 5: Evaluation using the Ebina et al. method.

	Precision	Recall	F-value
5 number analysis	56.55%	50.53%	53.38%
10 number analysis	57.98%	67.02%	62.17%
15 number analysis	58.49%	53.72%	56.01%

## 6 CONCLUSION

Many data are necessary to judge deflation in real time. Few tweets have geotag “coordinates” that can specify the position. Therefore, we use machine learning to estimate the location of tweets that have a geotag “place”, which is ambiguous location information for each municipality. Deflation was inferred from tweets that were presumed to have been posted at a specific location. The Endo et al. method did not improve the judgment accuracy, but the Ebina et al. method did.

In other words, SVM location estimation was able to increase the number of tweets posted at the target event location. However, because the accuracy of the position estimation is not sufficient, when judging the movement of people using the obtained tweets, the accuracy of the judgment may decrease as opposed to increasing.

Although SVM was adopted for machine learning, various other methods for machine learning are available. Results obtained from using them must be compared in future studies. In addition, it is necessary to consider methods other than judging the movement of people as a method of using the tweet whose position is estimated. There are various target events that can be considered when judging the movement of people in the same way. These will also be future research.

## 7 ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 19H01744.

## REFERENCES

- [1] T. Sakaki and Y. Matsuo, Twitter as a Social Sensor: Can Social Sensors Exceed Physical Sensors? *Journal of Japanese Society for Artificial Intelligence* 27(1), pp.67-74 (2012).
- [2] J. Kleinberg, Bursty and hierarchical structure in streams, In *Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.91-101, ACM(2002).
- [3] Y. Zhu and D.Shasha, Efficient Elastic Burst Detection in Data Streams, *Proc. 9th ACM SIGMOD International Conference on Knowledge Discovery and Data Mining*, pp.336-345, ACM(2003).
- [4] D. Shasha and Y. Zhu, *High Performance Discovery in Time Series, Techniques and Case Studies (Monographs in Computer Science)*, Springer-Verlag(2004).
- [5] X. Zhang and D. Shasha, Better Burst Detection, *Proc. 22nd International Conference on Data Engineering*, pp.146-149, IEEE computer Society(2006).
- [6] R. Ebina, K. Nakamura and S. Oyanagi, A Proposal for a Real-Time Burst Detection Method, *DBSJ Journal*, Vol.9, No.2, pp.1-6(2010).
- [7] R. Ebina, K. Nakamura and S. Oyanagi, A Proposal for a Real-time Burst Analysis Method, *IPSIJ TOD*, Vol.5(3), pp.86-96(2012).
- [8] M. Endo, Y. Shoji, M. Hirota, S. Ohno and H. Ishikawa, On Best Time Estimation Method for Phenological Observations Using Geotagged Tweets, *IWIN2016*, pp.205-210 (2016).
- [9] M. Endo, M. Hirota, S. Ohno and H. Ishikawa, Best-Time Estimation Method by Region and Tourist Spot using Information Interpolation, *IWIN2016*, pp.209-216 (2017).
- [10] O. Ozer, R. Heri and N. Kjetil: “Locality-adapted Kernel Densities for Tweet Localization”, *SIGIR'18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 1149-1152(2018).
- [11] T. Morikuni, M. Yoshida, M. Okabe, and K. Umemura: "Geo-location Estimation of Tweets with Stop Words Detection", *Journal of Information Processing, Transaction on Database* Vol.8, No.4, pp.16–26(2015).

# Anomaly detection in FA equipment using an interaction model

Hiroaki Ando<sup>†</sup>, Yusuke Iwatsuki<sup>†</sup>, Daiki Hibi<sup>†</sup>, Kazuhiko Tsutsui<sup>‡</sup>, Satoshi Aoki<sup>‡</sup>,  
Katsuhiro Naito<sup>†</sup>, Naoya Chujo<sup>†</sup>, Tadanori Mizuno<sup>†</sup>, and Katsuhiko Kaji<sup>†</sup>

<sup>†</sup>Aichi Institute of Technology, Japan

<sup>‡</sup>Mitsubishi Electric Corporation, Japan  
b20704bb@aitech.ac.jp

**Abstract** - Machine learning is being investigated as a promising method for detecting anomalies. Studies have proposed a method for detecting anomalies in equipment by searching for data that identifies equipment anomalies by arranging sensors in the equipment. However, this method requires a large number of sensors and is, therefore, expensive to purchase and install. In this research, we introduce a less costly method that does not use a large number of sensors. Factory Automation (FA) involves a lot of equipment and, when in operation, the machines are affected by the interaction between this equipment, specifically, the sounds, vibrations, and heat. We believe it is possible to detect equipment anomalies and reduce cost by modeling the vibrations caused by this interaction when the equipment is in its normal state and comparing it with its anomalous state. In this study, we collected normal and anomaly data, modeled interaction, and conducted evaluation experiments. We used the coefficient of determination for anomaly detection to compare the normal data model with the anomaly data.

## 1 INTRODUCTION

Factory Automation (FA) is the automation of factory production. Because FA operates over extended periods of time, malfunctions can occur. Such equipment malfunctions include situations where small screws and manufactured parts loosen and fall between equipment, making it difficult for the machine to move, and the expected wear from continued use.

Machine learning has been studied as a method for detecting equipment anomalies[1][2]. These studies place a large number of sensors in the equipment to search for and collect data that detects equipment anomalies. However, because this method requires many sensors, the purchase of these sensors and the labor required to install them is expensive. This method is also difficult to implement because the huge amount of data acquired from the sensors must be processed. In this research, we use the vibration caused by the operation of the equipment to detect equipment anomalies.[3][4][5][6]. An associated research generated an electromagnetic wave waveform from a physical model to diagnose anomalies in concrete structures[7]. An associated research has built physical characteristics and anomaly time models for various anomaly detection[8]. In our research, we generate an interaction model from a normal vibration waveform to detect anomalies due to the interaction of the equipment.

Our aim is to establish a method of anomaly detection that does not require a large number of sensors. FA involves multiple pieces of equipment resulting in interaction between the

equipment; that is, when one piece is in operation, the generated noise, vibration, and heat affects the other equipment. We focus on this interaction and only attach sensors to the affected side. Generally, vibration is the result of equipment interaction. Thus, the smaller the vibration, the more normal the equipment's state is. In this research, if an equipment's state is smaller than the expected vibration due to the interaction of other equipment, it is an anomaly. By collecting and modeling vibrations due to interaction under normal conditions from the sensor data, we believe that anomalies can be detected by comparing the data collected and modeled from the equipment.

Modeling sensor data when the equipment is in a normal state makes it easier to detect anomalies. Currently, we rely on heuristics to detect anomalies. If there is an anomaly, visualizing and judging the anomaly is possible without special knowledge or skill, and we can concentrate on solving more difficult problems and improving anomaly detection technology.

An additional advantage is cost reduction. Since current anomaly detection methods use a large number of sensors, purchase and installation costs are high. Developing a model will reduce this cost. For example, if four production machines are lined up in one production line, they can interact with each other. If sensors are placed on the first and third machines in consideration of their influence, the second and fourth machine can be installed without sensors because we can grasp the status of the second and fourth machines from the system's sensors. Building an anomaly detection system involves users, developers, designers, operation managers, and more. As the scale of the system to be built increases, more people will be involved and the costs will increase. By using modeling, the system will be built based on the model, fewer personnel will be required to build the system, and overall cost will be reduced.

## 2 ANOMALY VIBRATION DETECTION DUE TO INTERACTION BETWEEN LINEAR MOTORS

### 2.1 Measurement of Vibration Due to Interaction

Figure 1 shows a schematic of the interaction of the equipment we used in this study, which is typical of equipment used in FA. When linear motor 1 is moved as shown in Fig.1(a), linear motor 2 moves slightly, as shown in Fig.1(b). There are mainly three vibration interactions. The first is the phe-

nomenon caused when one machine stops and vibrates slightly due to the vibration of the operating equipment. The second is where two pieces of equipment influence each other's vibrations when both operate in synchronization. The third is when the pieces operate asynchronously and affect each other by vibration. We focus on the interaction of vibrations during simple operation in which a stopped machine vibrates slightly due to the vibration of the equipment still in operation. The model is currently designed to consider only under damping, but we are considering applying the model to oscillation and resonance in the future. Resonance is a phenomenon that occurs due to the interaction of equipment vibrations. Resonance is where vibration increases greatly when another vibration is added to it. We believe that we must study interactions due to vibration from simple motion to complex motion. In this research, we first study a modeling method for simple motion and then consider a modeling method for complex motion. If simple interactions can be modeled, we expect that the combination of simple interactions can explain complex interactions. For example, using the model, we are considering applications such as predicting and warning of resonant behavior setting.

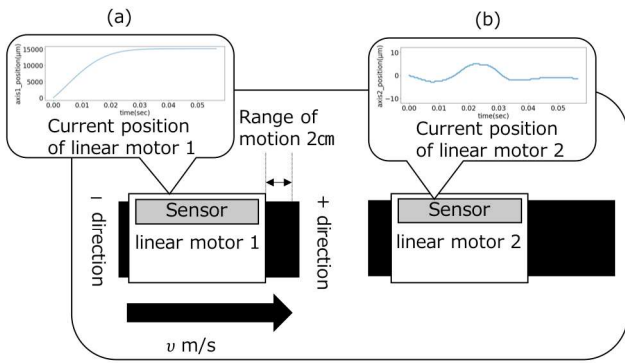


Figure 1: Schematic diagram of interaction measurement

## 2.2 Modeling of Interaction Data

We measured the normal data as shown in Fig.2 and identified both an amplification zone and a decay zone. The amplification zone is the zone where the vibration is amplified and the decaying zone is the zone where the vibration decays. Since the vibration of linear motor 2 tends to amplify in the amplification zone and decay in the decay zone, we created separate models for each zone.

We reduced the parameters to be set and modeled the interaction between the equipment due to vibration using sin and log functions. As shown in Fig.3, the sin function is suitable for expressing the waveform due to vibration. We created our model by assuming that the multiplication of the log function can express the amplification of the amplification zone and the decay of the decay zone.

A smoothing region is provided to prevent discontinuity in the slope if the amplification and decay intervals are simply connected. Smoothing is performed in the region of  $\Delta t$  seconds before and after centered on  $t_1$ . Spline interpolation is

used for smoothing.  $\Delta t$  is set as the time width necessary to perform smoothing. In this case, sufficient smoothing was achieved when  $\Delta t$  was set to 15 samples before and after the spline interpolation. For cubic polynomials on each small section that are separated by a finite number of equidistant points, a cubic polynomial whose derivative up to the second order is continuous (third-order) is called a spline. The interpolation condition for the given  $n$  data  $(x_k, y_k)$  ( $k = 1, \dots, n$ ) is

$$M(x_k) = y_k (k = 1, \dots, n).$$

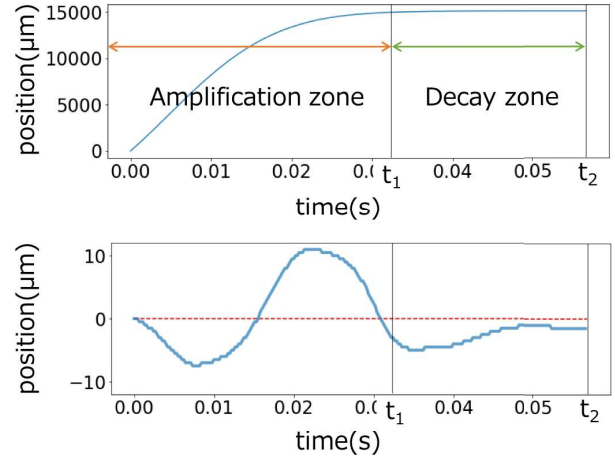


Figure 2: Amplification zone and decay zone

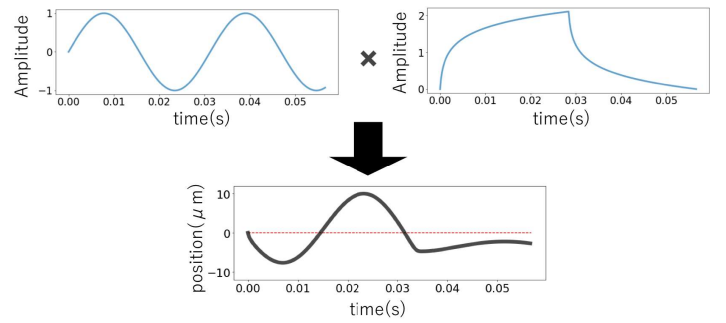


Figure 3: Composite function to represent vibration

Let the sin function formula be  $s(t)$  Eq.(1) and the log function formula be  $u(t)$  Eq.(2). Let  $t$  be time, and  $0$  to  $t_1$  be the amplification zone and  $t_1$  to  $t_2$  be the decay zone. We apply spline interpolation to the zone from  $(t_1 - \Delta t)$  to  $(t_1 + \Delta t)$  Eq.(3).  $\phi_1$  and  $\phi_2$  are phases, and  $\phi_1$  is a parameter that determines the phase that sets whether the movement onset moves to the positive side or the negative side when the vibration is expressed by the sin function.  $\phi_2$  is a parameter that determines the phase set to connect the amplification zone and the decay zone seamlessly.  $f_1$  and  $f_2$  are oscillation frequencies,  $a_1$  and  $a_2$  are constants for setting the magnitude of amplitude, and  $o$  is a constant for adjusting the start position of amplitude.



$$s(t) = \begin{cases} \sin(2\pi f_1 t + \phi_1) & (0 \leq t \leq t_1) \\ \sin(2\pi f_2 t + \phi_2) & (t_1 < t \leq t_2) \end{cases} \quad (1)$$

$$u(t) = \begin{cases} a_1 \log_e t & (0 \leq t \leq t_1) \\ a_2 \log_e t & (t_1 < t \leq t_2) \end{cases} \quad (2)$$

The model  $v_t$  is created by multiplying the sin function shown above and the logarithmic function. The model formula is shown below.

$$v(t) = \begin{cases} s(t)u(t) + o & (t < t_1 - \Delta t, t_1 + \Delta t < t) \\ M(s(t)u(t) + o) & (t_1 - \Delta t \leq t \leq t_1 + \Delta t) \end{cases} \quad (3)$$

### 2.3 Optimal Parameter Search Using GridSearch

GridSearch is used to improve the accuracy of the model. This method tries all parameter combinations. An image is looped, all combinations of parameters are tested, and the one with the best evaluation accuracy is searched. In our research, frequency are  $f_1, f_2$ , constants are  $a_1, a_2$  for setting the amplitude magnitude, the amplitude start position is  $o$ , and phase are  $\phi_1, \phi_2$ . GridSearch is applied to all parameters to find the optimal parameter.

However, the amount of calculation is enormous because GridSearch tries all combinations of the specified parameters. Therefore, in order to reduce the amount of calculation as much as possible, parameters are hit by measuring the frequency band using fast Fourier transform (FFT). We believe the calculation amount can also be reduced by performing a rough search followed by an additional detailed search.

### 2.4 Anomaly Detection

Anomaly detection is performed by comparing the generated model with the collected anomaly data to see what differences and characteristics they have. We validate the model by comparing the collected normal data with the model. Typical methods for evaluating the validity of a model include mean square error, root mean square error, and coefficient of determination. We evaluated the degree of agreement between the generated model equation and the collected normal data using the coefficient of determination ( $R^2$ ), which represents simple performance of the model. The reason we do not use anomaly detection algorithms such as one-class SVM is that we do not apply the model only to anomaly detection. We don't apply the anomaly detection algorithm because we are even thinking about the movement that causes oscillation and resonance. The closer the coefficient of determination is to 1, the higher the accuracy and the better the agreement with the model formula. In addition, the coefficient of determination is not expressed as 0 to 100, as with the ratio. Therefore, if the model formula and collected data are completely different, a negative parameter may be obtained. We consider the validity of the model from that parameter. Then, we compare the model confirmed to be valid with the anomaly data using the coefficient of determination. We believe that comparing

the difference between the model and the anomaly data will lead to anomaly detection in the equipment.

The standard deviation  $\sigma$  in the normal distribution is a numerical parameter of the size of the variation; after obtaining all deviations of individual data, we take the average and then take the average deviation of the measurement data as the average parameter of the variation. For the area within the range of standard deviation  $\pm \sigma$ ,  $\pm 2\sigma$ , and  $\pm 3\sigma$ , the probability of how much measurement data exists for the total (100%) area of the histogram of the normal distribution is the average parameter  $\mu$  as a center. Each standard deviation has the same probability. In other words, it is the same regardless of the unit. These probabilities are obtained from the standard normal distribution  $N(0, 1^2)$ . For this study, the threshold for anomaly detection is  $\mu - 3\sigma$ . The probability that there is data in the range of standard deviation  $\pm 3\sigma$  is about 99.7%. In this study, parameters greater than  $+3\sigma$  from the average parameter  $\mu$  are normal data, so parameters smaller than  $-3\sigma$  are anomalies.

## 3 EVALUATION EXPERIMENT

Figure 4 illustrates our approach. First, one linear motor (hereinafter referred to as a motor) is operated, the other motor is stopped, and the motor position data is collected. The vibration of the moving motor causes a slight vibration to the stopped motor. The data is collected using vibration as the normal data for the interactions. An interaction model (hereinafter "model") is created from the ordinary data and the vibration of a stopped motor is limited by weights or springs. If the movement on the stop side is limited, vibrations smaller than the normal vibrations will be obtained as data. If the vibration is smaller than normal, it may be an anomaly. This is the anomaly data. Otherwise, data is collected under the same conditions as collecting normal data. The model created from the normal data is compared with the anomaly data and used for equipment anomaly detection.

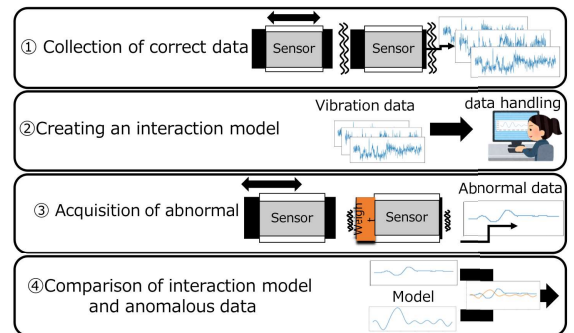


Figure 4: Research approach

### 3.1 Normal Data Collection

We collected data for modeling the interaction of equipment as physical properties such as vibration and friction. Vibration, sound, heat, etc. can be considered as factors that bring about the interaction; for the purposes of our research,

we focus on the physical interference of vibration, which is easy to verify, and then model.

Our experimental equipment includes two motors to measure the interaction, as shown in Fig.5. We affixed the two motors to the base and operated the motors via commands from the control software. The current position information of each motor was collected by the control software. In addition, the sensor data is possible collect even if there is no motion command.

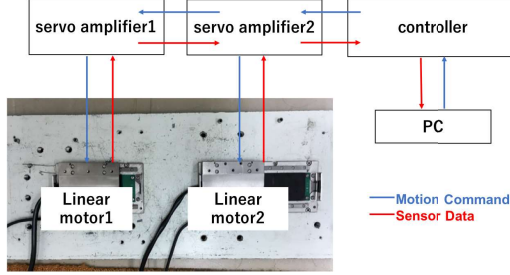


Figure 5: Configuration diagram of measuring equipment and experimental linear motors

We first moved one linear motor and measured the interaction of the other linear motor. In the next step, we placed an object on the linear motor to be operated or an anomaly was detected from the interaction when both linear motors move. When both move, the vibration may be amplified or damped, so it is necessary to generate a model to evaluate it. Ultimately, our goal is to introduce this as a system for detecting anomalies in FA equipment.

In this experiment, we used the equipment in Fig.1 to acquire the normal vibration data required for model creation. As shown in Fig.1(a), motor 1 is moved once from left to right at any speed, and we measured the vibration caused by the moved motor 1 to motor 2 when motor 2 was not moving. As shown in Fig.1(b), a slight movement was observed at the current position of motor 2 in the stopped state. We performed this experiment 10 times, collected the normal data and, based on the normal data, created a model by the method outlined below. Figure 6 shows the data when measured 10 times at 1.5m/s.

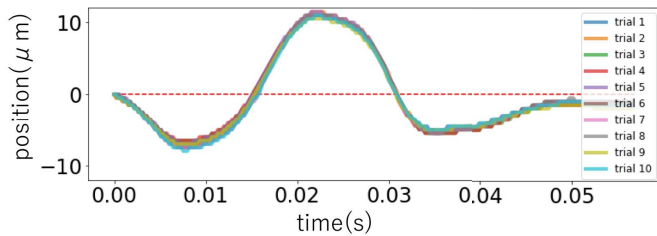


Figure 6: Normal vibration data collected

### 3.2 Vibration Modeling

We generated an interaction model based on the measured normal data. The concrete parameters of the model are 0.0340s

for  $t_1$ , 0.0568s for  $t_2$ , and -2 for  $\phi$ . In the amplification zone, the phase  $\phi_1$  is  $7\pi/6$ , the frequency  $f_1$  is 29 Hz, and the constant  $a_1$  is 22. The phase  $\phi_2$  is  $11\pi/6$ , the frequency  $f_2$  is 20 Hz, and the constant  $a_2$  is 5 in the decay zone. Table 1 summarizes the parameters.

Table 1: Optimal parameter obtained by GridSearch

	$t_1(s)$	$t_2(s)$	$\phi_1$	$\phi_2$	$f_1(Hz)$	$f_2(Hz)$	$a_1$	$a_2$	$\phi$	$R^2$
Rough	0.0340	0.0568	$7/6\pi$	$3/2\pi$	28	20	22	4	-3	0.928
Detailed	0.0340	0.0568	$7/6\pi$	$11/6\pi$	29	20	22	5	-2	0.954

GridSearch tries all combinations of the specified parameters at runtime, which is computationally expensive. Therefore, to reduce the calculation amount as much as possible, we measured the frequency band by FFT (See Fig.7). We also believe that we can reduce the amount of calculation by performing a rough search with GridSearch followed by an additional fine search. Table 1 shows the specific parameters searched by GridSearch. The parameter of the determination coefficient is larger in the detailed search than in the rough search (See Fig.8).

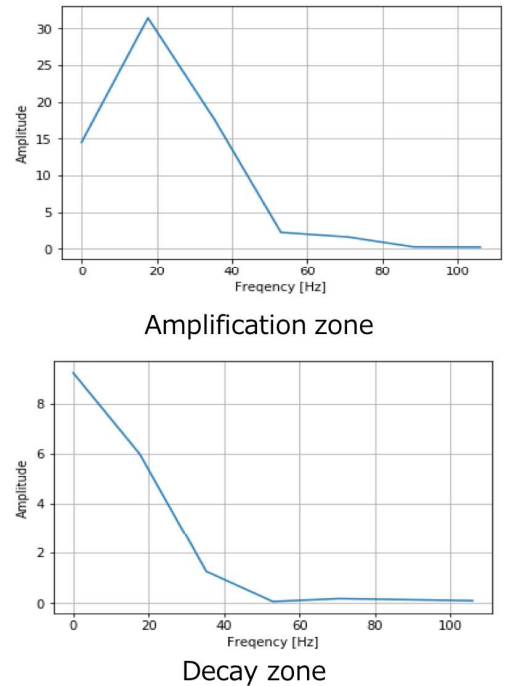


Figure 7: Frequency band measurement using FFT

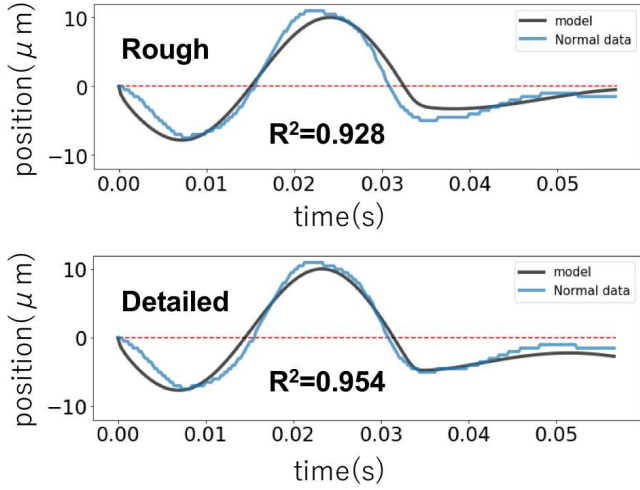


Figure 8: Difference between rough model and detailed model

### 3.3 Anomaly Data Collection

We collected the anomaly data to compare with the model. Motor anomalies include situations where small parts such as screws are mixed on the rail, making it difficult to move, as well as typical wear due to long-term operation. Since FA operates for long continuous time periods, the equipment must be durable and not prone to breaking down. The equipment containing the two motors used as our experimental equipment is difficult to break. Therefore, when collecting anomaly data, we need to assume an anomaly occurrence and reproduce it.

To reproduce the situation where small parts such as screws are mixed in the rail, thus hindering movement, we installed an obstacle on the rail to restrict the motor's movement (See Fig.9). Figure 10 shows the measured anomaly data. Other methods of inhibiting the motor's movement include installing a spring at the tip of the motor and placing a weight on the motor. We confirmed that vibration was suppressed by overlaying the normal and anomaly data. We investigated whether it is possible to detect anomalies using these data.

### 3.4 Anomaly Detection

We compared the model with the normal and anomaly data measured 10 times each at 1.3, 1.4, and 1.5m/s (See Fig.11). The coefficient of determination for comparing normal data with the model was 0.95 and the coefficient of determination for comparing anomaly data with the model was -3.55. The coefficient of determination for anomalous data is well below the threshold. We plotted the parameter of the coefficient of determination when compared with the model, as shown in Fig.12. The blue points are normal data, black points are anomaly data, and red lines are the threshold. Table 2 summarizes whether normal data and anomaly data can be classified by threshold parameters. As a result, we classified 100% of the normal data as normal and 100% of the anomaly data as anomaly. Anomaly detection is therefore possible because the

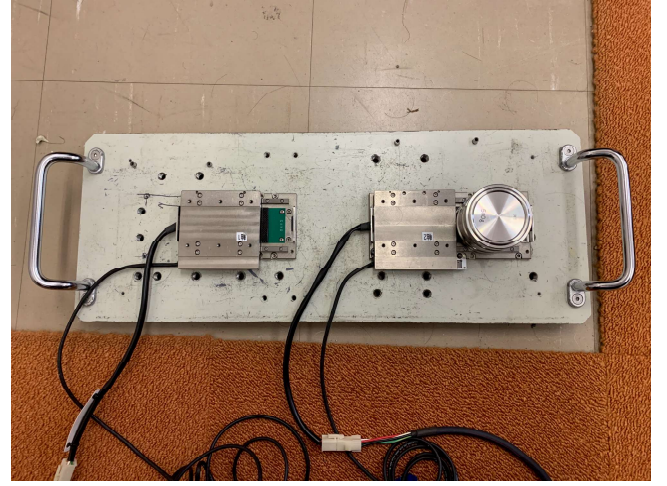


Figure 9: Experimental environment for collecting abnormal data

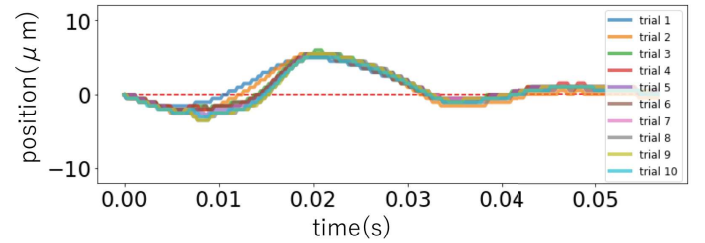


Figure 10: Collected pseudo-anomaly vibration data

model and the anomaly data are not close.

Table 2: Anomaly detection accuracy evaluation

		Estimation	
		Normal	Anomaly
Ground truth	Normal data	100%	0%
	Anomaly data	0%	100%

## 4 CONCLUSION

In this research, we focused on vibrations, which are relatively easy to verify and whose phenomena can be readily confirmed. Our experimental setup included two linear motors from which we measured the interaction data due to vibrations from the motor position information while changing the speed. We modeled the collected sensor data and compared the model with the data collected under normal conditions. When collecting anomaly data to be compared with the model, we installed obstacles on the rails to inhibit motor movement in order to reproduce anomalies. We used the coefficient of determination to compare the collected data. We compared the normal data with the model to see how well the model reproduces normal vibration. We then compared the model with the anomaly data. As a result of this comparison, we were able to classify the normal data as 100% normal and



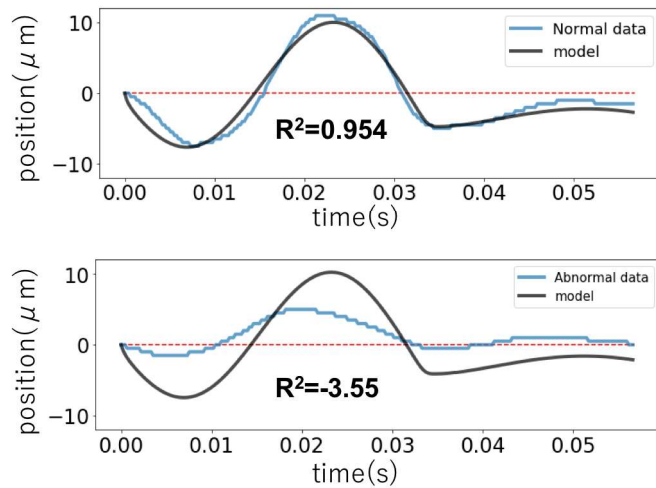


Figure 11: Comparison of model and anomaly data

the anomaly data as 100% anomaly. Since the coefficient of determination is closer to 1 as it approximates, we can conclude that the model and the anomaly data are not close to each other and that anomaly detection is, therefore, possible.

At present, velocity between 1.4 and 1.5m/s is not known without basic data, so we are trying to make it possible to succeed even without basic data. In this study, we searched for the optimum parameter using GridSearch, but it takes about 20 minutes to finish the program because GridSearch tries every combination one at a time. In the future, instead of trying all combinations, we are considering using another method that requires less computation. The number of data used to create the model this time was 10 which is a small amount of data. We plan to use more data in the future. The response to vibration should be modeled like the impulse response, but this time the model changes every time, so it is impossible, but we will have to consider it in the future.

## REFERENCES

- [1] U. K. Gupta, M. Ali, "Hierarchical representation and machine learning from faulty jet engine behavioral examples to detect real time abnormal conditions", IEA/AIE'88: Proceedings of the 1st international conference on Industrial and engineering applications of artificial intelligence and expert systems, Volume 2, pp.710 1988.
- [2] T. Oba, K. Yamada, N. Okada, K. Tanifuji, "Condition Monitoring for Shinkansen Bogies Based on Vibration Analysis", Journal of Mechanical Systems for Transportation and Logistics, Vol.2, Issue 2, pp.133-144 2009.
- [3] S. Liu, F. G. A. Ball, "Detection of engine valve faults by vibration signals measured on the cylinder head", Proceedings of the institution of Mechanical Engineers Part D Journal of Automobile Engineering, Volume: 220, issue: 3, pp.379-386 2006.
- [4] W. Zhaoxia, L. Fen, Y. Shujuan, W. Bin, "Motor Fault Diagnosis Based on the Vibration Signal Testing and

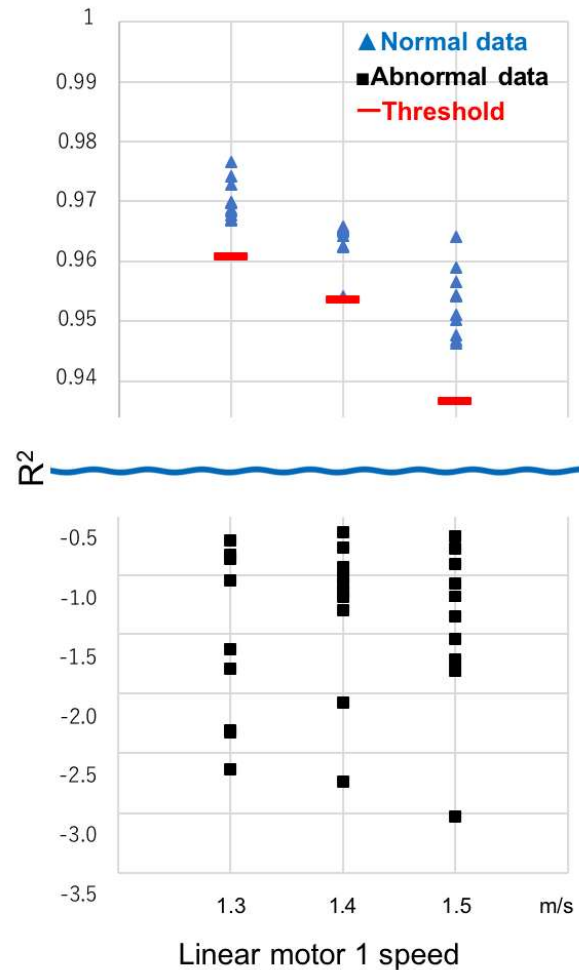


Figure 12: Classification as normal or anomaly

- Analysis", Third International Symposium on Intelligent Information Technology Application 2009.
- [5] Q. Sun, T. Cao, Y. Hou, T. Zhao, "Detection and Analysis Based on the Abnormal Mechanical Vibration Signal of GIS", Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC) 2015.
- [6] I. S. Koo, W. W. Kim, "The development of reactor coolant pump vibration monitoring and a diagnostic system in the nuclear power plant", ISA Transactions 39 pp.309-316. 2000.
- [7] S. Tanaka, M. Yamada, "Non-Destructive Inspection of Concrete Structures Using an Electromagnetic Wave (Radar) Based on a Signal Propagation Model", Transactions of the Society of instrument and Control Engineers", Volume 39, Issue 5, pp.432-440, 2003.
- [8] J. J. Gertler, "Survey of model-based failure detection and isolation in complex plants", IEEE Control Systems Magazine, Volume 8, Issue 6, Dec. 1988.



# Object Extraction Method from Mobile Camera Videos Using Optical Flow

Tsukasa Kudo<sup>†</sup>

<sup>†</sup>Faculty of Informatics, Shizuoka Institute of Science and Technology, Japan  
kudo.tsukasa@sist.ac.jp

**Abstract** - Nowadays, image recognition using deep learning is rapidly developing, and object and its state recognition are performed from images in various fields. On the other hand, with the progress of IoT, it has become possible to easily collect target images at various places by using videos of mobile cameras such as wearable cameras. Here, when the ratio of the target in the image is small, the recognition accuracy deteriorates due to the influence of the background. So, firstly, the target area extraction from the image is often performed as a preprocessing, and image recognition is performed by using this area. However, since the mobile camera itself moves, it is difficult to extract this area in conventional ways. In this paper, I propose a method to extract the target area from the video frames. The frames are superimposed with matching their background positions by using optical flow, then the target is extracted based on the difference between frames. And, I show the proposed method can extract the target area even in front of complex backgrounds.

**Keywords:** Optical flow, Image processing, Object detection, Mobile camera, Wearable camera, Video

## 1 INTRODUCTION

In recent years, the effectiveness of deep learning in image recognition has been shown, and its use in various fields is progressing [4], [9], [24]. For example, handwritten character recognition and face recognition are widely performed by the multi-class classification model of deep learning [1], [5]. However, when applying deep learning, collecting enormous training data often becomes a problem.

On the other hand, with the progress of the Internet of Things (IoT), various sensors are connected to the network and a large amount of data is collected and accumulated. Regarding videos, a large number of cameras have been deployed and are used for various purposes such as monitoring the traffic conditions and inside of buildings. And, image recognition using deep learning is widely applied to automatically identify objects and detect abnormalities from such a large amount of videos [16].

For such a study, I attempted to discriminate the location and target using deep learning from images extracted from videos of mobile cameras such as wearable cameras. As a result, I showed that the training data could be collected efficiently and the discrimination accuracy could be improved by continuous discrimination against the same target while comparing the results [11]. On the other hand, I found when the target is small, there was the problem of deterioration of the accuracy due to the influence of the background.

To suppress the influence of the background, the preprocessing of the image for deep learning is often performed, in

which a relatively small area including the target is extracted from the image. In this study, we target backgrounds that are various kinds but do not move such as indoor walls and equipment. Also, since it targets mobile cameras, it is possible to shoot a movie while moving the target in the hand in front of the camera.

To extract the target from the frame of the video, there are the background subtraction method, the frame subscription method, and utilization of optical flow. In the background subtraction method, the target is extracted by the difference between the background image and the image in which the target appears in front of the background; in the frame subscription method, the target is extracted by its difference in each frame. However, in both cases, it is assumed that the background image is fixed, and it has been pointed out that it is difficult to apply it directly to a moving camera such as a mobile camera [15].

And, for a video, the optical flow shows the displacement vector between a pixel of one frame and the corresponding one of another frame. And in the dense optical flow, this applies to all pixels [7]. When the target held in a hand is moved in front of the mobile camera, the area of the target in the optical flow can be extracted as a largely moved part. However, for the mobile camera, since it also moves, the background area of the video also moves. So, firstly, I show there is a problem that the complicated background affects the optical flow and disturbs extracting the target.

In this study, for this problem, I propose a method to extract the target from various backgrounds utilizing the optical flow, in which the backgrounds of two frames are superimposed with matching their background positions based on the optical flow then the difference of frame is generated. As a result, the difference of the target area is obtained, because the motion of the target is different from the background. Therefore, the target is extracted by using this difference. Furthermore, it is shown through experiments that the target area can be extracted with high accuracy when the targets in both frames have a certain displacement distance.

The remainder of this paper is organized as follows. Section 2 shows the challenges of optical flow in a complicated background and the motivation of this study, and Sec. 3 proposes a target area extraction method based on the optical flow. Section 4 shows the implementation of this method in the experimental system, and Sec. 5 shows the experimental results and its analysis. Section 6 concludes this paper.

## 2 MOTIVATION AND RELATED WORKS

I have been working on improving inventory management work in a machine factory, where various parts are stored in

a bulk container. Since these inventory quantities cannot be counted visually from the outside, its stock-taking becomes a heavy workload for workers. For this problem, I showed the inventory quantity can be estimated with practical accuracy from the image of the bulk container by applying deep learning [10]. However, since there are usually more than 1000 bulk containers in a factory, an efficient collection way of these images became the next challenge.

And for this challenge, focusing on the fact that inventory fluctuations occur when workers replenish and ship the parts, I conceived to estimate stocks using images extracted from videos of wearable cameras worn by workers. To estimate the inventory quantity from only the video, firstly it is necessary to detect the movement to the bulk container and the work on the parts. For the former, I showed it was possible to detect it with a certain accuracy by using the deep learning model trained to distinguish the entrance and equipment of the target room by using the images automatically extracted from videos [11].

For the latter, I noticed that the worker needs to hold the parts in his hand for the inventory work. In other words, as shown in Fig. 1 (a), if the object held in the hand can be recognized, the target parts can be distinguished automatically. And, I collected various such images to train the model, and evaluate the accuracy of distinguishing the target. As a result, I found that there is a problem that the accuracy deteriorates for small targets because of the influence of the background.

For such problems, the method is usually adopted in which an area including the target is extracted from the image, then the target is recognized by using the area image. For example, in face recognition, the face area in an image is firstly extracted by using such as Harr-like features, then face is recognized with this area [25]. So, firstly, I conducted an experiment in which the target was moved in front of the wearable camera to extract the target based on the difference in the optical flow between the background and target. As a result, the target could be extracted with high accuracy when the background was flat. However, I found there is a problem in the case shown in Fig. 1 (a) where the background was complicated and the target was flat. The target area was divided as shown in Fig. 1 (b), and it was difficult to extract the target area as a continuous region.

The process of this extraction is as follows. Figure 1 (c) is a visualization of the displacement direction of the optical flow, and the brightness increases counterclockwise from 0 (black) that shows the right direction. Figure 1 (d) is a visualization of the normalized displacement distance, in which the higher the brightness is, the larger the displacement distance is. As shown in (c) and (d), the brightness distribution was not separated between the background and target. Figure 1 (e) is a binarization of the brightness in (c), in which the white area corresponds to the hand and the upper part of the book. Similarly, in Fig. 1 (f), the white area corresponds to the same or more brightness area than the upper part of the book in (d). These white areas are the area including the target estimated from the optical flow, and (b) was created by superimposed the original image on the union of the white area in (d) and (e).

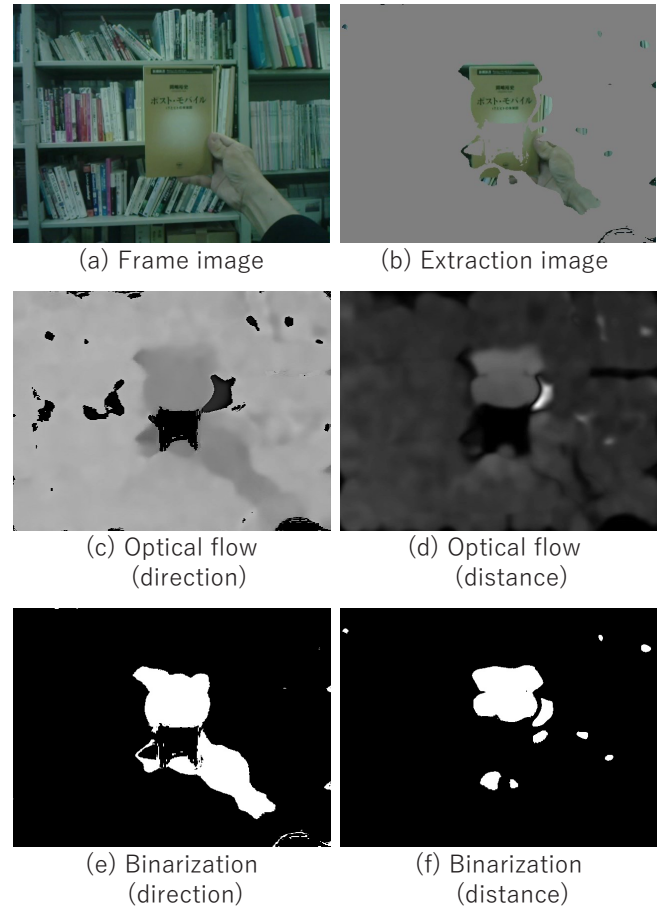


Figure 1: Problem of target extraction with optical flow

The motivation of this study is to develop a method that can accurately extract the area including the target, even in the case against such a complex background with a free-moved mobile camera. Besides, this study targets the stationary background and the rigid objects that can be held and moved by the hand as shown in Fig. 1 (a). Also, it is necessary to be able to execute efficiently because the images of the target part are continuously extracted from the moving image, although there are some intervals.

In recent years, in addition to wearable cameras as in this study, many kinds of mobile cameras have widespread such as in-vehicle cameras and mobile phone cameras. So, many studies have been conducted to detect and track a moving object from the videos of mobile cameras.

Regarding the object detection and classification of detected objects, researches utilizing deep learning have been progressing rapidly. Faster R-CNN performed both of them in a lump by collective end-to-end training of both models [20], and YOLO executed them by a single neural network to improve efficiency [19]. Then, for different scale objects, SSD enabled to process them collectively [14], and RetinaNet improved efficiency by introducing the Feature Pyramid Network (FPN) and improving the error function [12], [13]. Furthermore, M2Det has further improved both accuracy and efficiency by introducing the new FPN and error function [29]. However, since these methods target each image, they are not suitable for object detection targeted by this study, which de-

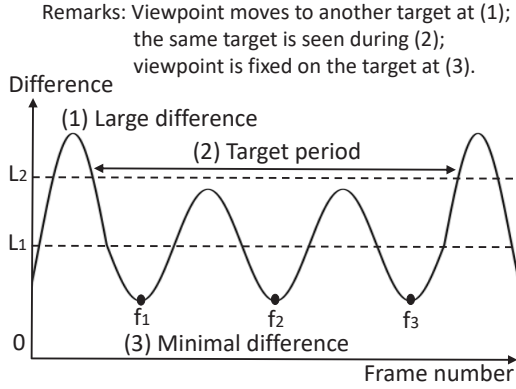


Figure 2: Transition of difference between adjacent video frames

tracks the objects moving in front of the background using multiple frames of the video. Furthermore, as a general problem, to apply these deep learning methods to the individual target shown in Fig. 1, it is necessary to prepare a large amount of training data and train the model.

On the other hand, for the free-moved camera, various studies using optical flow have been conducted [27]. The most direct ways use the distance difference in the optical flow, which happens between the background and the moving target [17], [26]. However, as shown in Fig. 1, it has been pointed out that this has a problem in the case of a complicated background.

For this problem, some methods have been proposed such as combining with other methods, performing analysis over many frames, and utilizing deep learning [2], [8], [15]. However, there are issues in terms of efficiency, such as processing complexity and model training. Also, the methods to estimate camera motion by utilizing optical flow have been proposed [6], [21]. However, these aims rather the motion recognition. Furthermore, some methods have been proposed in which the background is reconstructed by utilizing the optical flow and the target is detected by the background subtraction [22], [28]. However, these target seamless backgrounds or pan-tilt-zoom cameras, that is, it is difficult to apply to the wearable camera shooting the complex background shown in Fig. 1.

To summarize the above, the efficient method has not been proposed to extract an area including a moving target from a complicated background with a free-moved camera.

### 3 PROPOSED METHOD

This study aims to extract the area including the target, namely the target area, from the frame of the video of the wearable camera as shown in Sec. 2 by utilizing the optical flow. First, to calculate the optical flow between frames, it is necessary to extract suitable frames. That is, it is necessary to extract frames with small blur from adjacent frame pairs that are looking at the same target. The proposed method determines them based on the difference between adjacent frames. Figure 2 shows an example of the transition of the difference between adjacent frames. (1) shows the point where the difference is large, that is, it is the case where the viewpoint is moving from one object to another. (2) shows the period where the difference is a certain value or less, that is, the case

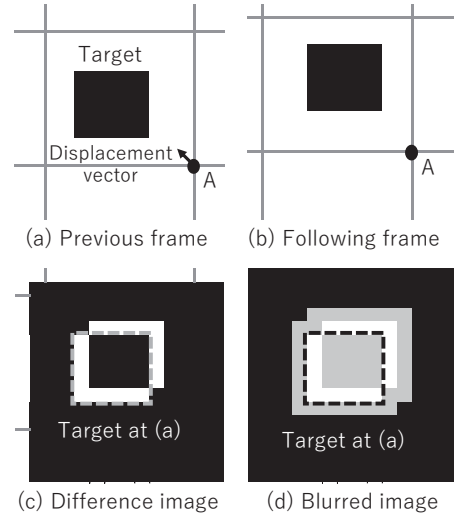


Figure 3: Target extraction method using optical flow and difference

where the same object is watched. (3) shows the point where the difference is minimal, that is, the blur of the frame image is small. Therefore, it is expected that the accuracy is improved by calculating the optical flow using the frame at point (3).

And, the target frame is extracted based on the thresholds  $L_1$  and  $L_2$  of the two levels as shown in Fig. 2.  $L_2$  is used to exclude frame pairs which difference is too large. The target frame is extracted from the “Target period” in Fig. 2 where the difference is less than  $L_2$ .  $L_1$  is used to exclude frame pairs which difference is too small, that is, the distance of their optical flow is too small. As a result, as shown by  $f_1$ ,  $f_2$ , and  $f_3$  in Fig. 2 as an example, the frame with the minimal difference is selected as the target frame in each period less than  $L_1$ .

Figure 3 shows the target area extraction process in the proposed method. (a) shows the previous frame of video such as  $f_1$ , and (b) shows the following frame such as  $f_2$ . Here, the black rectangle is the target. In this study, as shown in Sec. 2, the background is assumed to be stationary, so the difference in the background between (a) and (b) is only the parallel translation. So, firstly, when the displacement vector between (a) and (b) is obtained for one point  $A$  in the background, (b) can be displaced so that the background of (a) and (b) superimposed by using this. As a result, their difference in the background area becomes 0, that is, black as shown in (c).

At this time, if the target was moved, there is a gap in the target area between two frames. And, this gap becomes the difference from the background as shown in the white area in Fig. 3 (c). Note that there is no difference in the central area of the gap where the target in both frames overlap, so it becomes black. Therefore, the target at (a) shown by the gray dashed rectangle is included in the gap including this central area. Here, a part of the difference area is outside of this rectangle. However, since this study aims to narrow down the area where the target exists, this area is acceptable.

Then, as shown in the gray area in Fig. 3 (d), the target area is enlarged by blurring the image; the entire interior of

the area is also targeted. Further, to exclude the different areas between frame (a) and (b) existing at the left end and the upper end in (c), this area is set as without difference (black). Finally, by extracting these white and gray areas as a continuous area, the area can be extracted, which includes the target at (a) indicated by the black dashed line in (d).

## 4 IMPLEMENTATION

To evaluate the proposed method, the functions described in Sec. 3 were implemented as an experimental system. I implemented it on a Windows 10 personal computer. For the programming language, Python Ver.3.6 was used; for the image processing opencv-python Ver.4.1.0.25 was used.

Besides, in this study, for comparative evaluation of the target extraction accuracy, we also implemented the target extraction function by using only the optical flow in addition to the proposed method.

### 4.1 Implementation by proposed method

Firstly, to extract target frames from video, the image of the frame is converted to grayscale, and a histogram of the number of pixels with brightness  $j$  is created. This number is expressed by  $n_{ij}$  by using the frame number  $i$  and brightness  $j$ . Then, as shown in Eq. (1), the absolute difference of  $D_i$  between previous and following frames is calculated by weighting with luminance  $j$  and dividing by the number of pixels ( $N$ ). Here, dividing is performed to keep  $D_i$  constant regardless of the number of pixels.

$$D_i = \sum_{j=0}^{255} |n_{i+1j} * j - n_{ij} * j| / N \quad (1)$$

Here,  $D_i$  corresponds to "Difference" in Fig. 2. As for the period while  $D_i$  is less than the threshold  $L_2$ , it is determined that the same target is watched. Among this period, for each section where  $D_i$  is less than the threshold  $L_1$ , the frame with the smallest  $D_i$  is extracted, such as  $f_1$  in Fig. 2.

For calculating the optical flow, calcOpticalFlowFarneback method of opencv-python was used [18]. This is an implementation of the polynomial expansion algorithm [3]. In this method, I set the parameter as follows: the polynomial area was 5, the polygon width was 0.5, the window size was 60, the pyramid size was 0.5, its level was 3, and Gaussian kernel was used for prior blurring. Figure 4 shows an example of intermediate results of processing by the proposed method. (1) shows the previous frame of the video, and the target book moves to the left. The background is almost flat. Similar to Fig. 1, (2) and (3) show the displacement direction and distance of the optical flow of (1). In this case, since the background is flat, the length of the displacement vector is almost zero as shown in (2) and (3). In this implementation, the optical flow of the background was visually checked and a point looks like average brightness was used to calculate the displacement of the background.

Next, as shown in (4), the difference between the previous frame (1) and its following frame was taken, and the distribution of luminance was expanded by histogram equaliza-

tion. Since the contrast was expanded by the histogram equalization, small differences in the background are also emphasized. (5) is an image obtained by blurring the image in (4) with a median filter, which was used for connecting the image fragments. And, fragments such as the title of the book are connected as shown in (5). (6) is a binarization image of (5) in which the area of the brightness above the threshold was extracted as white parts. As the threshold, 159 was used, which is the median value 127 plus the error 32. As a result, as shown in (6), the part corresponding to the target was extracted as a continuous area by binarization. Also, similar to Fig. 3 (d), along with binarization, a certain range from the edge of the frame is made black as a non-target area to delete the non-overlapping part of two frames made by frame displacement.

And, as shown in (7), the image of the target could be displayed in the target area after the following processing was performed on (6). First, erosion and dilation of the white area were performed to exclude noises and to separate unnecessarily combined regions. Each of them was performed three times with a kernel (3,3). Second, specifying the start point, which is the point in the largest area among areas that are overlapping the target, the target area was extracted as a continuous area. In this implementation, the start point was specified manually. Third, the entire part surrounded by the convex hull contour was extracted, then the corresponding part of the previous frame was displayed there. Fourth, the range out of the target area was painted in gray, which was the pre-process of the next blur processing.

Lastly, as shown in (8), the contour blurring of (7) was performed. Since the proposed method aims to recognize the target by using deep learning, this is intended to reduce the effect of the boundary edge. Gaussian blurs with kernel size 127, 73, 31, 15, 7, and 3 were performed sequentially, in which the target area was replaced to the original image before each blur to maintain the target image. By this procedure, the contours of both the boundary and the target area were smoothly blurred as shown in (8).

### 4.2 Implementation by using only optical flow

The process of the target extraction by using only the optical flow is the same as the proposed method until to create the optical flow image shown in (2) and (3) of Fig. 4. Then, the range of brightness including the target image brightness was specified, and the area of this range was binarized as the area including the target as shown in Figs. 5 (1) and (2). This range was determined manually by referencing the histograms in the brightness of the two optical flow displacement images. Next, as shown in (3), a union of the two white areas in (1) and (2) was created, and a convex hull contour was created similar to Fig. 4 (7). (4) shows the image in which the original frame was displayed on this convex hull contour. Then, the contour of the target area in (4) was blurred similar to Fig. 4 (8).

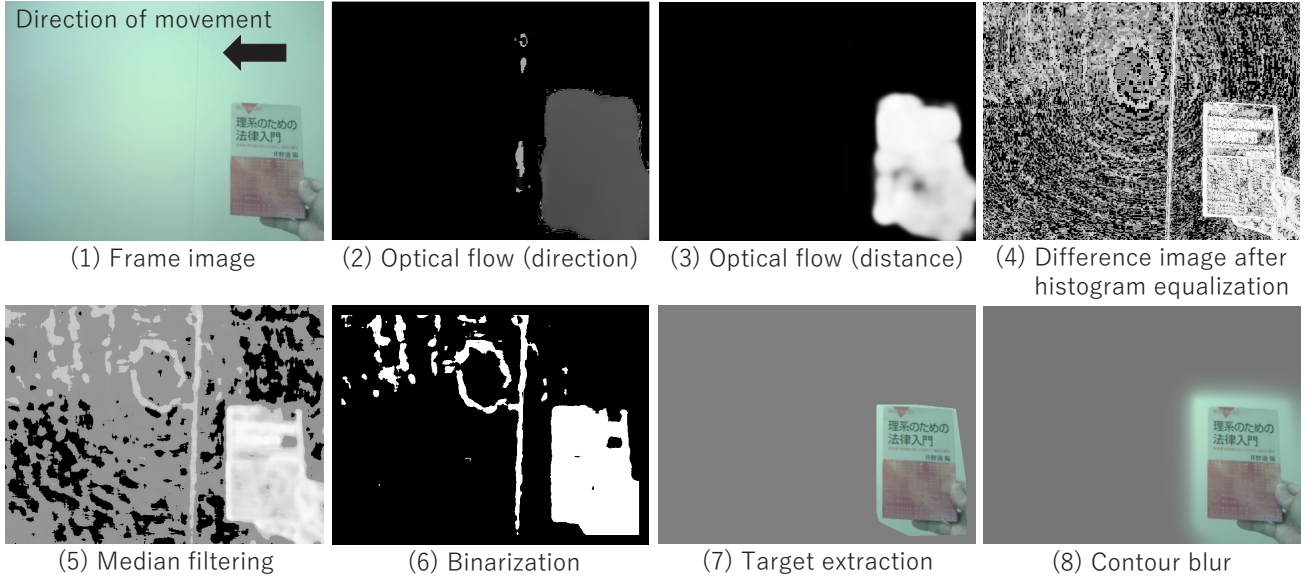


Figure 4: Intermediate processing result images in the proposed method

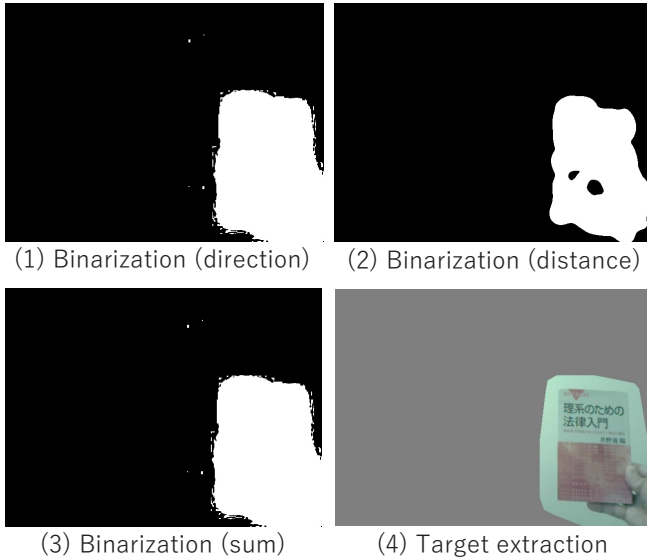


Figure 5: Target containing area based on optical flow

## 5 EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the effectiveness of the proposed method, comparative evaluations of extraction accuracy were conducted using two extraction methods implemented in Sec. 4, that is, both the proposed method and the method by using only the optical flow.

### 5.1 Experimental Environment

The three types of targets, namely books, shown in Table 1 and the four types of backgrounds shown in Table 2 were used as the experimental environment, and the accuracy was evaluated with images combining them. As shown in Fig. 6, books consisted of the following: B1 had a clear outline in the lower half, B2 had a clear figure though did not have a

Table 1: Target object in experiment

No.	Type	Used target
B1	Clear contour	Book 1
B2	Clear figure	Book 2
B3	Flat	Book 3

Table 2: Background object in experiment

No.	Type	Used backbroung
W1	Flat	Wall without equipment
W2	Sparse	Wall with equipment
W3	Bordered	Wall with large monitor
W4	Complex	Book shelf

clear outline, and B3 had a relatively flat image. As shown in Fig. 6, backgrounds consisted of the following: W1 was a flat wall; W2 was a wall with equipment placed in front and relatively simple; W3 was a background with a clear boundary by the monitor; W4 was a complicated background of the bookshelf.

These images were taken by a wearable camera at almost the same time in a laboratory. For the wearable camera, head-set EPSON MOVERIO Pro BT-2000B shown in Fig. 7 was used. it equipped a video camera on the part that hits the forehead as shown by the arrow in Fig. 7 and displayed its video on the glasses. It was used with a frame size of 640 x 480 dpi and 30 frames per second. Images were extracted from the video shot by this camera by the experimental system mentioned in Sec. 4.



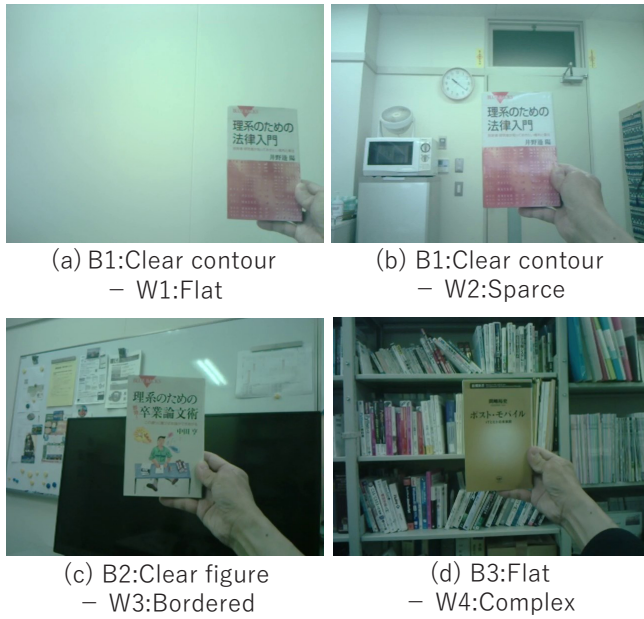


Figure 6: Target and background in experiment

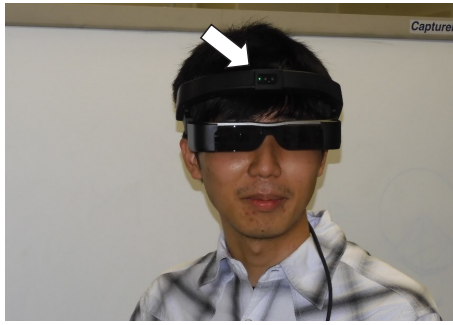


Figure 7: Wearable camera used in experiment

## 5.2 Evaluations with Adjacent Frames

First, using the difference between frames calculated by Eq. (1), adjacent five frames were extracted to create adjacent four pair frames such as the pair  $(f_1, f_2)$ ,  $(f_2, f_3)$  as shown in Fig. 2. And, four experiments were conducted for each combination of a target and a background by using them. Figure 8 shows the transition of the difference fluctuation in the frames of book B1. The relatively flat periods in Fig. 8 corresponds to the one in which the book was moved in front of each background in Table 2; the relatively large fluctuation corresponds to the displacement from one background to another one. Note that the magnitude of these fluctuations is different among backgrounds, and it is larger in W4 (complex) than in the others. So, in this experiment, the threshold  $L_1$  in Fig. 2 was set to 2.0 for W2 and 1.0 for the others.

Next, for all combinations of the three books and the four types of backgrounds, comparative evaluation of the extraction accuracy of the target by the proposed method and by using only the optical flow was conducted. Figure 9 shows the results. The accuracy was set to 1.0 when the target can be extracted with the convex hull contour shown in Fig. 4 (7) or in Fig. 5 (4); it was set to 0.5 when though the target is

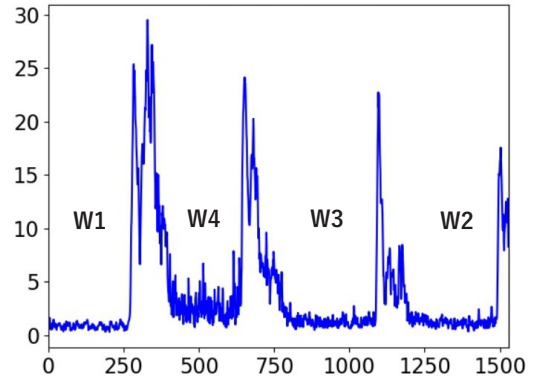


Figure 8: Transition of difference from previous frame

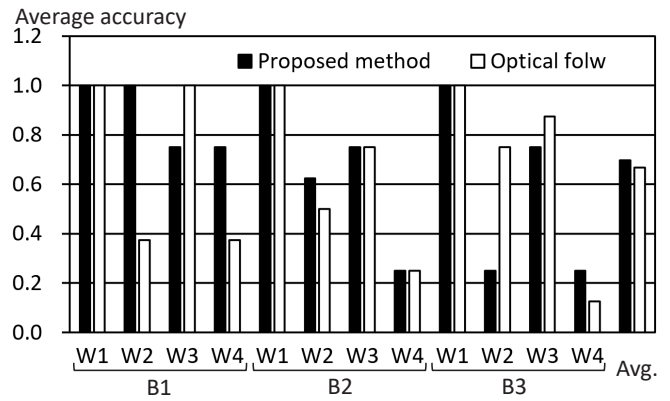


Figure 9: The average accuracy of target extraction with adjacent frames

not completely extracted in this case but completely extracted with the contour blur such as Fig. 4 (8); it was set to 0.0 when the target is not completely extracted even with the contour blur. And, the average on the above-mentioned 4 pair frames was calculated.

As shown in Fig. 9, for B1 (Book 1), the proposed method could achieve relatively high accuracy. Besides, though a part was missing after the convex hull contour processing in the case of the backgrounds W3 (Bordered) and W4 (Complex), it was completely extracted after the contour blur. However, for B2 and B3 (Books 2 and 3), the fluctuations in the accuracy were large, and there was no significant difference from the method by using only the optical flow as shown “Avg” (average) at the right end.

Furthermore, for B1 in Fig. 9, the accuracy was lower than by using only the optical flow only in the case of the background W3. So, its cause was analyzed. Figure 10 (1) shows the displacement distance of the optical flow and Fig. 11 shows its histogram. As shown in Fig. 10 (1), though there was a bright part in the upper right from the target, the target area was almost captured. Here, as shown in Fig. 11, the maximum displacement distance is 2.8 pixels, which is the displacement distance in the upper right bright area. However, the brightness difference between the background and target was equivalent to 0.6 pixels, in which the former’s brightness is 0 (black). This shows that since the target’s displacement distance was extremely small, a clear difference image of the

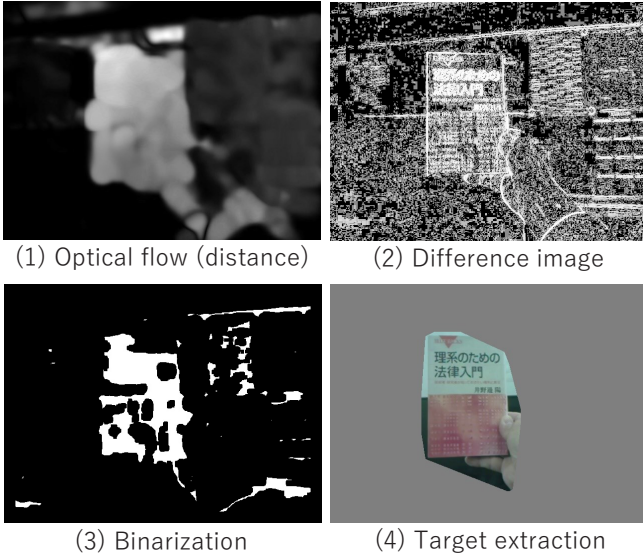


Figure 10: Analysis of image missing part of target

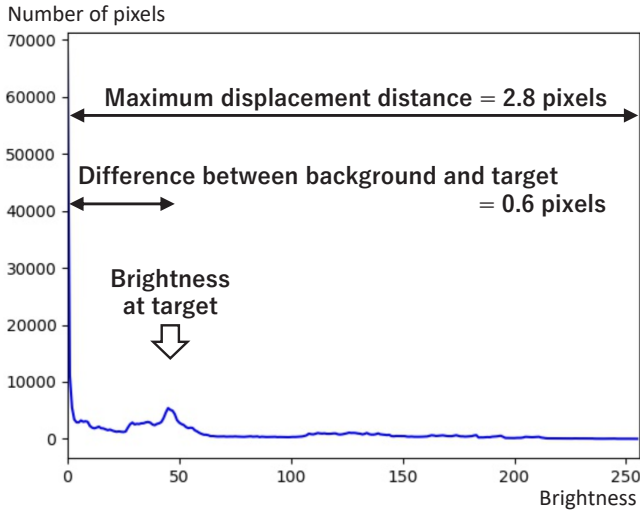


Figure 11: Histogram of optical flow of distance

target could not be obtained as shown in Fig. 10 (2). As a result, as shown in Fig. 10 (3), the upper right of the target was missing while the binarization; it was also the same in extraction target in convex hull contour as shown in Fig. 10 (4).

### 5.3 Evaluations with Increasing Frame Interval

Based on the analysis results in Sec. 5.2, I evaluated the change in the accuracy of the proposed method while the frame interval was increased for objects B2 and B3. In this experiment, frame  $f_1$  was used for the previous frame in all the case, and the frame intervals were sequentially extended as of  $(f_1, f_2)$ ,  $(f_1, f_3)$ . The same frame was used as in Sec. 5.2.

Figure 12 shows the result. The horizontal axis shows the number of frame intervals, and the vertical axis shows the average accuracy of each target against four backgrounds in

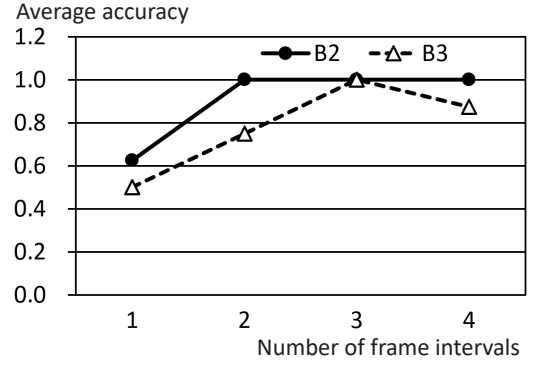


Figure 12: Transition of average accuracy with number of frame intervals

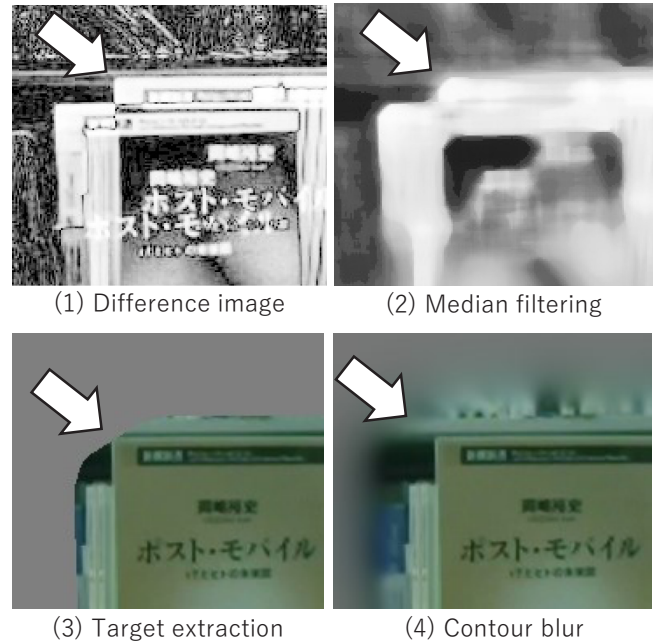


Figure 13: Analysis of missing vertices

Table 2. The accuracy was improved by extending the frame interval, and when the interval was 3, the average accuracy became 1.0 in both targets. However, in the case of interval 4, for target B3 and background W4, the corner was missing while the convex hull contour although it could be completely extracted after the contour blur. The previous frame in this case was the one shown in Fig. 1 (a).

I analyzed this case and found that although the target was extracted in the difference image as shown in Fig. 13 (1), the corners were rounded while the median filter processing as shown in (2). As a result, the corner of the target was missing in the extraction by the convex hull contour as shown in (3). However, this missing was so small that it was completely extracted by the contour blur processing as shown in (4).

This method assumes blurring to use for deep learning, so the final area expands. Therefore, this problem does not occur while this method is performed with blurring. However, when the image after the convex hull contour is used, the corner should be also extracted in this image. For this purpose, for example, in the erosion and dilation processing mentioned in

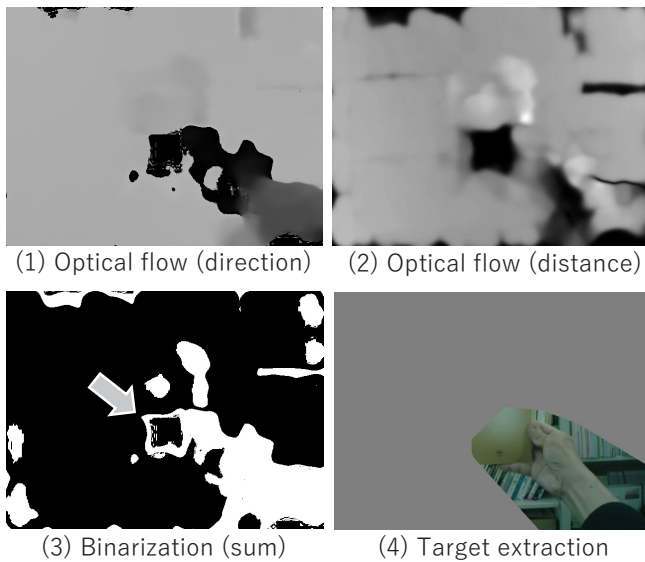


Figure 14: Analysis of extraction failure with optical flow

Sec. 4.1 and shown in Fig. 4 (7), the target area should be dilated enough to recover the rounding by the filter.

Furthermore, the method by using only optical flow was evaluated in the same case namely the target B3 and the background W4. As a result, the accuracy was 0.0 in all number of frame intervals. Figure 14 shows the case where the number of frame intervals was four. As shown in (1) and (2), in this case, unlike the case shown in Fig. 5, a continuous area including the target could not be obtained. So, similar to Fig. 1 (b), it was divided into multiple areas while the binarization as shown in Fig. 14 (3). Therefore, as shown in Fig. 14 (4), the whole target area could not be extracted even after the convex hull contour. Note that the upper left part of the quadrangle indicated by the gray arrow with white frame in Fig. 14 (3) was separated as noise while the erosion processing and excluded.

## 5.4 Evaluation Results

As shown in Figs. 12 and 13, when using an image after the contour blur, the area including the whole target could be extracted by the proposed method. However, it was necessary to adjust the interval between the using frames appropriately. Here, as shown in Fig. 11, the max displacement distance of pixels between frames can be grasped from the optical flow. Therefore, it is considered that an appropriate frame pair can be obtained by excluding the frames if this distance is not in the designated certain range.

Besides, when the target area extracted by the proposed method was used without the contour blur, I found the corners might be missing as shown in Fig. 13. So, as mentioned in Sec 5.3, the dilation should be performed according to the rounding by the median filter.

Furthermore, the motivation for this study is to collect the data for training and discrimination for deep learning efficiently. For this purpose, it is necessary to automate the selection of the following points mentioned in Sec. 4.1: the representative background point in the optical flow, and the area

including the target after the binarization. These are challenges of the future study along with the automation of the above-mentioned extracting the appropriate frame pairs.

## 6 CONCLUSION

By using videos from mobile cameras such as wearable cameras, we can efficiently collect data for training and discrimination for deep learning. However, for a small target, since the discrimination accuracy deteriorates due to the influence of the background, it is necessary to extract a relatively small area including the target from the video frame. However, for the mobile cameras, there was a problem in accuracy to extract the target from various backgrounds in the frame image by the conventional method.

For this problem, I proposed a method to extract the target by creating a difference target image between the frames, in which the backgrounds are superimposed based on the displacement vector in the optical flow. Furthermore, through the experiments, it was shown that the target area can be extracted with high accuracy by selecting the frame pair having the appropriate displacement distance of the target.

The future study will focus on the automation of the proposed method to efficiently extract the images of the target area from a large amount of video.

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 19K11985.

## REFERENCES

- [1] F. Chollet, "Deep learning with python," Manning Publications Co. (2017).
- [2] A. Elqursh, and A. Elgammal, "Online moving camera background subtraction," European Conference on Computer Vision, pp. 228–241 (2012).
- [3] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," Scandinavian conference on Image analysis, Springer, pp. 363–370 (2003).
- [4] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, "Deep learning," MIT press (2016).
- [5] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, and T. Hospedales, "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," Proc. IEEE Int. Conf. on computer vision workshops, pp. 142–150 (2015).
- [6] M. Jain, H. Jegou, and P. Bouthemy, "Better exploiting motion for better action recognition," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2555–2562 (2013).
- [7] A. Kaehler, and G. Bradski, "Learning OpenCV 3: computer vision in C++ with the OpenCV library," O'Reilly Media, Inc. (2016).
- [8] A. I. Károly, R. N. Elek, T. Haidegger, K. Széll, and P. Galambos, "Optical flow-based segmentation of moving objects for mobile robot navigation using pre-trained



- deep learning models,” 2019 IEEE Int. Conf. on Systems, Man and Cybernetics, pp. 3080–3086 (2019).
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, pp. 1097–1105 (2012).
- [10] T. Kudo, and R. Takimoto, “CG Utilization for Creation of Regression Model Training Data in Deep Learning,” *Procedia Computer Science*, Vol.159, pp. 832–841 (2019).
- [11] T. Kudo, “A Proposal for Article Management Method Using Wearable Camera,” *Procedia Computer Science* (in press).
- [12] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2117–2125 (2017).
- [13] Lin, T. Y., P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” *Proc. IEEE Int. Conf. on Computer Vision*, pp. 2980–2988.(2017).
- [14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” *European Conf. on Computer Vision*, pp. 21–37, Springer. (2016).
- [15] K. Makino, et al., “Moving-object detection method for moving cameras by merging background subtraction and optical flow methods,” *2017 IEEE Global Conf. on Signal and Information Processing*, pp. 383–387 (2017).
- [16] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, “Deep learning for IoT big data and streaming analytics: A survey,” *IEEE Communications Surveys & Tutorials*, Vol. 20, No. 4, pp. 2923–2960 (2018).
- [17] M. Narayana, A. Hanson, and E. Learned-Miller, “Coherent motion segmentation in moving camera videos using optical flow orientations,” *Proc. IEEE Int. Conf. on Computer Vision*, pp. 1577–1584 (2013).
- [18] OpenCV team, “OpenCV modules,” <https://docs.opencv.org/4.3.0/index.html> (reffered May 25, 2020).
- [19] J. Redmon, S.Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 779–788.(2016).
- [20] S. Ren, K. He, R. Girshick, and J. Sun., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, *Advances in Neural Information Processing Systems*, pp. 91–99 (2015)..
- [21] S. Singh, C. Arora, and C. V. Jawahar, “Trajectory aligned features for first person action recognition,” *Pattern Recognition*, Vol. 62, 45-55 (2017).
- [22] M. Unger, M. Asbach, and P. Hosten, “Enhanced background subtraction using global motion compensation and mosaicking,” *2008 15th IEEE Int. Conf. on Image Processing*, pp. 2708–2711 (2008).
- [23] Y. Nagai, H. Masuta, T. Motoyoshi, K. Sawai, T. Tamamoto, K. I. Koyanagi, and T. Oshima, “A study of Optical Flow in Peripheral Vision Area while Vehicle Cornering,” *2018 Joint 10th Int. Conf. on Soft Computing and Intelligent Systems and 19th Int. Symposium on Advanced Intelligent Systems*, pp. 1327–1331). IEEE. (2018).
- [24] M. Verhelst, and B. Moons, B., “Embedded deep neural network processing: Algorithmic and processor techniques bring deep learning to iot and edge devices,” *IEEE Solid-State Circuits Magazine Vol.9, No. 4*, pp. 55–65 (2017).
- [25] P. Viola, and M. Jones, “Rapid object detection using a boosted cascade of simple features,” *Proc. 2001 IEEE computer society conference on computer vision and pattern recognition*, Vol. 1, pp.511-518 (2001).
- [26] H. Wang, H. and C. Schmid, “Action recognition with improved trajectories,” *Proc. IEEE Int. Conf. on Computer Vision*, pp. 3551–3558 (2013).
- [27] M. Yazdi, and T. Bouwmans, “New trends on moving object detection in video images captured by a moving camera: A survey,” *Computer Science Review*, Vol. 28, pp. 157–177 (2018).
- [28] W. Zhang, X. Sun, and Q. Yu, “Moving Object Detection under a Moving Camera via Background Orientation Reconstruction,” *Sensors*, Vol. 20, Issue 11, 3103 (2020).
- [29] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling, “M2det: A single-shot object detector based on multi-level feature pyramid network,” *Proc. AAAI Conf. on Artificial Intelligence*, Vol. 33, pp. 9259–9266.(2019).



# Automatic detection of tourist spots and best-time estimation using social network services

Munenori Takahashi<sup>\*</sup>, Masaki Endo<sup>\*\*</sup>, Shigeyoshi Ohno<sup>\*\*</sup>,  
Masaharu Hirota<sup>\*\*\*</sup>, and Hiroshi Ishikawa<sup>\*\*\*\*</sup>

<sup>\*</sup> Electronic Information Course, Polytechnic University, Japan

<sup>\*\*</sup> Division of Core, Polytechnic University, Japan

<sup>\*\*\*</sup> Faculty of Informatics, Okayama University, Japan

<sup>\*\*\*\*</sup> Graduate School of System Design, Tokyo Metropolitan University, Japan

{m20305, endou, ohno}@uitech.ac.jp

hirota@mis.ous.ac.jp

ishikawa-hiroshi@tmu.ac.jp

**Abstract** – As described in this paper, we investigate the establishment of a method that requires no prior knowledge of tourist attractions to estimate the optimal viewing time for cherry blossoms. Geotagged tweets are useful as a social indicator for estimating and acquiring local tourist information in real time because the information from tweets can reflect real-world situations. Earlier studies of methods of estimating cherry blossom viewing times have typically relied on the assumption that one knows a tourist destination: it is impossible to estimate cherry blossom tourist spots that a system user does not know. As described in this paper, we attempted automatic detection of spots using geotagged tweets with heat maps and visualization using configuration conditions. We described results obtained from verifying the optimal time to detect the auto-detected spot.

**Keywords:** Mining, Sightseeing, SNS, Spot detection

## 1. INTRODUCTION

In recent years, opportunities are increasing for tourists to obtain tourism information using the web. Particularly because of the spread of social networking services (SNSs), various information is distributed and accumulated on the web. Some SNSs, such as Twitter, can accumulate and disseminate location information. We are currently using information from Twitter's geotagged tweets to estimate the best time to view cherry blossoms. A low-cost method [2] using a moving average was proposed to use this information for estimating the optimal time for observing phenologic phenomena. The proposed method can estimate the best time to view cherry blossoms in prefectures and municipalities where tweets with geotags are sufficiently obtainable. Moreover, the geotagged tweets used for this method are useful as a social indicator elucidating real-world situations. This method of estimating viewing times can use local tourism information effectively in real time.

Existing methods require that users have a known and preferred destination for cherry blossom viewing. Therefore, existing methods can not provide information related to the best time to see cherry blossom spots at locations that are unknown to users. For users to be comfortable using the system, it must be independent of the user's knowledge. Therefore, this study assesses a method for automatic spot

detection to be used in combination with existing methods to estimate optimal timing. This method establishes a comprehensive extraction method that is useful for tourists who want to make phenological observations: they can detect locations and estimate the best cherry blossom viewing times.

As described herein, we conducted an experiment to ascertain whether cherry blossom sites can be detected automatically from tweets, or not. This report explains that the automatic spot detection can reveal unknown spots.

## 2. RELATED RESEARCH

Diverse information such as location information, images, and character strings are accumulated continuously and in large amounts by SNSs. Earlier research efforts have examined extraction of such information from SNSs.

Yoshida et al. [3] predicted a tendency for tweets to be retweeted frequently. The tendency was examined using two methods: Potentially Retweeted Scoring (PWRTS) and Reverse retweet propensity scoring (IRTS). Experiment results have shown that the IRTS trend for Tweet detection might be worth considering in both English and Japanese.

Isokawa et al. [4] detected changes in local regional characteristics based on patterns from earlier statistical congestion data. They analyzed changes while particularly addressing regional characteristics. Using statistical analyses for which a spatial mesh and change points were set in two counties, they confirmed regional changes quantitatively when a specific new, large facility was created.

Hubert et al. [5] performed sentiment analysis by finding correlation in public reactions to government tweets. They demonstrated identification of people's emotions in response to government tweets and visualized the emotions.

Guangyao et al. [6] described that the core content improves a category-based collaborative filtering algorithm based on the user authority. The newly proposed algorithms overcome shortcomings of traditional collaborative filtering, i.e., the problem of an extreme sparsity matrix.

Maenaka et al. [7] proposed the Sakura Sensor, a participatory sensing system that extracts landscape route information automatically from videos taken using an in-vehicle smartphone. It then shares the data among user nearby in near real time. Using the method described by Maenaka et

al., we were able to confirm cherry blossoms in a flowering state with accuracy of about 74% and a recall rate of 84%. In addition, the k-stage sensing method achieved the same point of interest detection rate in half the sensing time as that shown by the conventional method.

Finally, although research on SNS and regional changes has been conducted as described above, no study of cherry blossom spot analysis has been reported. Therefore, for this study, we detected the cherry blossom sightseeing spots automatically and estimated the best time to view them.

### 3. PROPOSED METHOD

For this research, we propose a method that can estimate an optimal viewing time irrespective of the amount of available knowledge. Earlier studies demonstrated that if a person can see the number of geotagged tweets in a prefecture or city, then that person can estimate the best time to visit. Estimation of the best time in Tokyo in 2018 by comparing the cherry blossom bloom date with the full bloom day [8] observed by the Japan Meteorological Agency demonstrates that the existing method [9] had 100% recall and 53.3% accuracy. It is noteworthy that the full-bloom day is a state in which about 80% or more buds of the sample tree are open. Therefore, the best time estimate continues even after the full-bloom day. Using this conventional method, one can estimate the best time for viewing after the full-bloom day for a prefecture or municipality unit. Furthermore, specification of the prefecture name and municipality is necessary at the time of estimation. For the map, the proposed method used the latitude and longitude information attached to the tweet. The number of tweets was visualized using a heat map. Subsequently, the conditions were set using the heat map. Tourist spots were detected. Using the existing method, we conducted an experiment to ascertain whether the automatically detected spots can be best-time estimated.

## 4. EXPERIMENT METHOD

This chapter presents an estimation method for automatic spot detection and for the best time to see it.

### 4.1 Preprocessing and Data Use

This section presents descriptions of preprocessing and the data used. Using the Streaming API [10], we collected geotagged tweets with location information including latitude and longitude in Japan. Then we analyzed tweets with biological names. For this experiment conducted during February 17, 2015 through July 29, 2019, the transition of tweets related to cherry blossoms was confirmed with analyzed organism names as "さくら", "サクラ", and "桜". For tweets in Tokyo, we used a simple reverse geocoding service [11] of the Research Institute of Agriculture, Forestry and Fisheries for latitude and longitude information and prefectures, municipalities, based on the latitude and longitude information of geotagged tweets that include location information. Analyses were conducted using general towns and streets within the name / city planning area. For tweets in Shizuoka, we extracted tweets with latitudes of approximately 34.603 to 35.416 and longitudes of

approximately 137.494 to 139.175. Analyses were conducted using latitude and longitude information.

### 4.2. Automatic Spot Detection

This section presents a description of the correct data and method used for automatic detection of cherry blossom spots. First, we set 64 correct answer spots corresponding to cherry blossoms in Tokyo from the Walker + 2019 edition of 1,000 views [12]. Similarly, we set 40 correct data for Shizuoka prefecture. Spots were determined when generating a 3D map using spreadsheet software (Excel; Microsoft Corp.) and using heat map shading to meet the conditions. Figure 1 presents the heat map shading and condition range. A small tweet quantity is displayed in black; large quantities are in white. There, range (1) is 85%–100%; range (2) is 60%–85% (0%, black; 100%, white).



Figure 1: Light and shade of the heat map and condition range.

The proposed method is a system for automatically detecting sightseeing spots: it is independent of the amount of user knowledge. At the initial stage, we investigate whether the proposed method can achieve automatic detection of existing spots. Therefore, we set a square with a 100 m side and a circle with 1 km diameter centered on the latitude and longitude indicated by Walker + as the correct answer spot. Then, three conditions were set as conditions for automatic detection: Condition 1 – the square contains tweet volume (1) (range (1) in Fig. 1); Condition 2 – the circle contains tweet volume (1) (range (1) in Fig. 1); Condition 3 – the circle tweet volume (2) (range of (2) presented in Fig. 1) is included. Using this condition, we confirm whether the correct answer spot can be detected automatically from the tweets. We show how many correct answer spots were detected using automatic spot detection. Then we calculate the spot automatic detection rate. The method obtains the value by dividing the spots detected automatically by the number of correct answer spots. Figure 2 portrays the squares and circles set within the correct range.

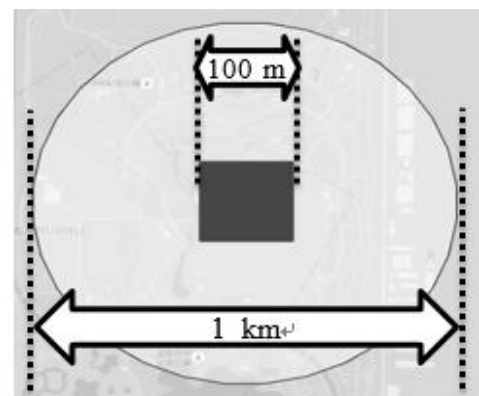


Figure 2: Range of correct answers.

### 4.3. Judgment of Cherry Blossom Viewing Time

As the best time estimation method, we use the existing method [8], which uses a weighted moving average. This section presents a description of existing methods.

### 4.4. Weighted Moving Average

The weighted moving average used for the existing method applies weights assigned to the respective values of data. By adding weights, we improved the recall and accuracy of best time estimation. With the existing method, the median was set to 1. Values  $\pm 0.5$  from the median were taken respectively as the minimum and maximum values. In addition, except for the median, weights from the lowest value to the highest value were assigned linearly. The third decimal place was rounded. Taking the 5-day weighted moving average used for the existing method as an example, the following equation (1) was obtained. Here,  $Have_5$  represents the weighted moving average for 5 days;  $x_y$  is the number of tweets  $x$  for  $y$  days prior.

$$Have_5 = (x_5 * 0.5 + x_4 * 0.75 + x_3 * 1 + x_2 * 1.25 + x_1 * 1.5) / 5 \quad (1)$$

### 4.5. Method for Best Time Estimation

Using a simple moving average and a weighted moving average, the following estimation method was set for the frequency of appearance of each geotagged tweet including the target word. Results were analyzed by date to estimate the best viewing period. 1) We used a one-year simple moving average to ascertain the period during which tweets about cherry blossoms increased. 2) Because the number of tweets tends to be higher on Saturdays and Sundays, the 7-day weighted moving average was used on a weekly basis. 3) A 5-day weighted moving average was used based on the average number of days from the flowering of cherry blossoms to full bloom, which is 5 days. The best time to see each tourist spot was estimated using these best time estimation criteria. Next, we set two conditions for the best time to see the cherry blossoms. Condition 1: Number of tweets  $\geq 1$ -year moving average, Condition 2: 5-days moving average  $\geq 7$ -days moving average And 3 consecutive days. The cherry blossoms are in full bloom when both conditions 1 and 2 are met.

## 5. EXPERIMENT RESULTS

This chapter presents results of automatic spot detection and its best time estimation.

### 5.1. Automatic Spot Detection

We visualize the tweet frequency by application of latitude and longitude information of geotagged tweets to the Excel heat map. An experiment was conducted to assess automatic spot detection using the conditions presented in 4.1 and the range of correct answer spots. As an example, Figure 3 shows spots (Jindai Botanical Park) that meet Condition 1. Figure 4 shows spots (Ukima Park) that meet Condition 2. Table 1 presents the same decision results for all 64 cases in Tokyo. Table 2 shows the spot automatic detection rates under conditions 1–3 obtained based on results of all 64 judgments in Tokyo. Table 3 presents the same decision results for all 40 cases in Shizuoka. Table 4 shows the spot automatic detection rates under conditions 1–3 obtained based on results of all 40 judgments in Shizuoka.

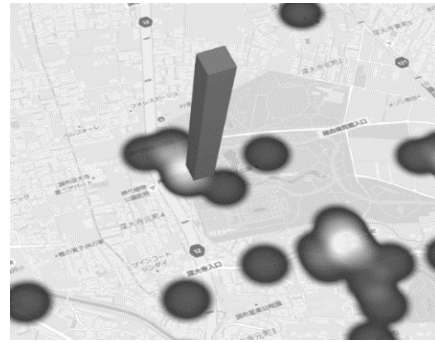


Figure 3: Correct answer spots that meet condition 1.

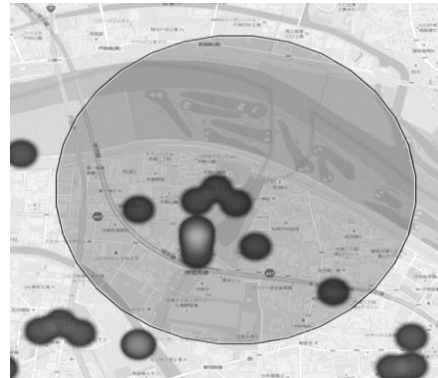


Figure 4: Correct answer spot that meets condition 2.

Table 1: Judgment results of all 64 correct answer spots in Tokyo

	Contains Gray below	Gray within 1km	Snow white within 1km		Contains Gray below	Gray within 1km	Snow white within 1km
ARK HILLS	0	0	0	CHIDORIGAFUCHI PARK	1	1	1
ASUKAYAMA PARK	1	1	1	CENTRAL PARK	0	0	0
SENZOKU POND PARK	0	0	1	TOKYO MIDTOWN	1	1	1
IKEGAMI HONMONJI TEMPLE	1	1	1	TOSHIMAEN SAKURA FESTIVAL	0	0	1
INOKASIRA GIFT PARK	1	1	1	TONERJ PARK	0	0	1
UENO GIFT PARK	1	1	1	TOYAMA PARK	1	1	1
UKIMA PARK	0	1	1	SAKURAGAOKA PARK	0	0	1
BAIGANJI	1	1	1	SAYAMA PARK	0	0	1
TAMAGAWADAI PARK	0	1	1	JINDAI BOTANICAL GARDEN	1	1	1
LAKE OKUTAMA	0	0	0	HIKARIGAOKA PARK	0	1	1
OTONASHISHINSUI PARK	1	1	1	YOYOGI PARK	1	1	1
ONE GREEN ROAD	0	1	1	NISHIARAI PARK	0	0	1
KAMANOFUCHI PARK	0	0	0	HIBIYA PARK	1	1	1
FORMER IWASAKI GARDEN	0	1	1	HAMARIKYU GARDENS	0	1	1
FORMER SHIBARIKYU GARDEN	1	1	1	HAMURA WEIR	1	1	1
FORMER FURUKAWA GARDEN	0	1	1	FUJIMORI PARK	1	1	1
KIYOSUMI GARDEN	0	1	1	HOMYOUT APPROACH	1	1	1
KOISHIKAWA KORAKUEN	0	1	1	MIZUMOTO PARK	0	0	0
KOGANEI PARK	0	1	1	MYOJINSHITA PARK	0	0	0
SHOWA KINEN PARK	0	0	1	MUKOJIMA HUNDRED GAEDEN	0	0	0
KOTTA RIVER	0	0	0	MUSASHINO PARK	0	0	1
KOMAZAWA OLYMPIC PARK	0	1	1	MEIJI JINGU GAIEN	0	1	1
SUNSHINE CITY	1	1	1	MEGURO RIVER	1	1	1
SHIOTAKOCHITSUTSUMI	0	0	0	ROPPONGH MOHRI GARDEN	1	1	1
SHIBA PARK	1	1	1	YAESU SAKURA STREET	1	1	1
SHAKUJI PARK	0	0	0	YOMIURI LAND	0	0	0
SHINJUKU GYOEN	1	1	1	RIKUGIEN	1	1	1
SUMIDA PARK	1	1	1	SOTOBORI PARK	1	1	1
SENDAIBORIGAWA PARK	0	0	0	KINUTA PAEK	0	0	1
ZENPUKUI RIVER GREEN	0	0	0	TATSUMINOMORI GREEN PARK	0	0	1
TAKIYAMA PARK	0	0	1	HARIMAZAKA	1	1	1
TAMA RIVER EMBANKMENT	0	0	1	YASUKUNI SHRINE	1	1	1

Table 2: Spot automatic detection rate in Tokyo

	Condition 1	Condition 2	Condition 3
Hit	27	39	51
Answer spot	64	64	64
Answer Spot automatic detection rate	42.2%	60.9%	79.7%

Table 3: Judgment results of all 40 correct answer spots in Shizuoka

	Contains Gray below	Gray within 1km	Snow white within 1km		Contains Gray below	Gray within 1km	Snow white within 1km
SHIZUOKA SENGEN SHRINE	0	0	1	MT IE	1	1	1
SUMPU CASTLE PARK	1	1	1	NOMORI POND	0	1	1
NIHONDAIRA	0	1	1	KIYA RIVER BANK	0	0	0
SHIMIZU FUNAKOSHITSUTSUMI PARK	0	0	1	KURIGATAKE	0	0	0
MT. GOTEN	0	0	1	TSUMAGOI RESORT SAINO SATO	0	1	1
HAMAMATSU CASTLE PARK	1	1	1	RENGE TEMPLE POND PARK	0	0	1
SANARU LAKE PARK	0	0	0	MT KONPIRA GREEN PARK	1	1	1
OKUYAMA PARK	0	0	1	FUJI PEACE PARK	0	0	1
AKIBA DAM	0	0	0	CHICHIBUNOMIYA MEMORIAL PARK	0	0	1
MT. TOBA PARK	0	0	0	KANO RIVER SAKURA PARK	0	1	1
CHITOSE RIVER	0	0	0	KATSUMATA RIVER	0	0	0
ITO RIVER PROMENADE	1	1	1	KAWAZU RIVER	1	1	1
ATAMI CASTLE	0	1	1	SHIMOGAMO HOT SPRING	1	1	1
MISHIMA TAISHA	1	1	1	MATSUZAKI TOWN	1	1	1
KARIYADO NO GEBAZAKURA	0	0	1	KOGANEZAKI PARK	0	0	0
TAISEKI TEMPLE	0	0	0	SURUGADAIRA NATUREL PARK	0	0	1
IZU KOUGEN	0	0	0	FUJI CEMETERY	0	0	0
SAKURA NO SATO	1	1	1	OKUNI SHRINE	0	0	0
ITO HOT SPRING	0	1	1	OTA RIVER SAKURA DANK	0	0	0
KAWARAMACHI OLD EMBANKMENT	0	0	0	SENGOKU YUMEKAIDO HIKING COURSE	0	0	0

Table 4: Spot automatic detection rate in Shizuoka

	Condition 1	Condition 2	Condition 3
Hit	10	16	25
Answer spot	40	40	40
Answer spot automatic detection rate	25.0%	40.0%	62.5%

As presented in Table 2, 27 correct answer spots were found for which the range of tweet volume (1) was included in the 100 m square of condition 1. The spot automatic detection rate was 42.2%. In condition 2, the number of correct answer spots in which the tweet volume (1) range was included in a circle of 1 km diameter was 39. The spot automatic detection rate was 60.9%. Under condition 3, the number of correct answer spots which included the range of tweet volume (2) within a circle with a diameter of 1 km was 51. The spot automatic detection rate was 79.7%.

As presented in Table 4, 10 correct answer spots were found for which the range of tweet volume (1) was included in the 100 m square of condition 1. The spot automatic detection rate was 25.0%. In condition 2, the number of correct answer spots in which the tweet volume (1) range was included in a circle of 1 km diameter was 16. The spot automatic detection rate was 40%. Under condition 3, the number of correct answer spots which included the range of tweet volume (2) within a circle with a diameter of 1 km was 25. The spot automatic detection rate was 62.5%.

The reason for the low spot automatic detection rate under condition 1 is that the range of the 100 m square was overly narrow. In fact, many spots are famous for cherry blossoms. Each one is a correct answer, such as a park or garden, but each covers a large area. A 100 m square cannot encompass such a large park. In Condition 2, the correct answer spot that was leaked in Condition 1 was also found. However, a new difficulty is that tweets are scattered to that extent in large parks and gardens. Because the tweet points of origin are not concentrated in one place, one can infer when regarding the tweet volume that the number of tweets did not increase to the threshold range of tweet volume (1). A large tourist spot such as a park or garden cannot be estimated correctly if the range of conditions is narrow.

Differences between results obtained in Tokyo and Shizuoka are explainable: an important difference is the number of tweets which can be extracted. About 145,000 tweets were extracted in Tokyo, but only about 14,000 were extracted in Shizuoka. The difference in the extracted tweets is therefore about 10 times: the number of spots that can be detected decreased automatically. However, it might be improved by increasing the number of data using other SNS. In Shizuoka Prefecture, correct answer spots are located at the mountains and at the foot of the mountains. Some of them are out of reach of radio waves. Some can only accommodate four buses a day. Therefore, moving by car is necessary. Perhaps tweeting is impossible because a person is driving. The tweet might not be detected because parking positions were sparse. Figures 5–7 present the respective communication areas, marked as gray, of three telephone services in Japan: au<sup>TM</sup>, NTT docomo<sup>TM</sup>, and Softbank<sup>TM</sup>.

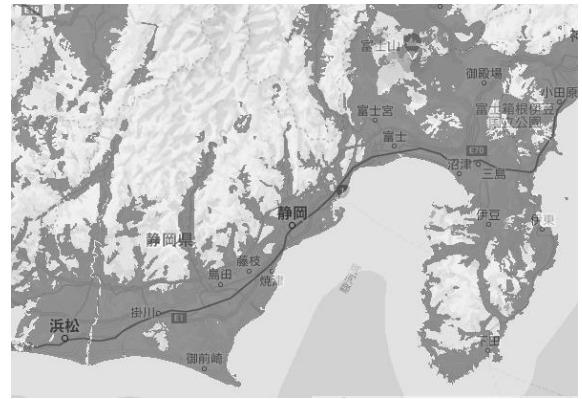


Figure 5: au<sup>TM</sup> communication areas.



Figure 6: NTT docomo<sup>TM</sup> communication areas.

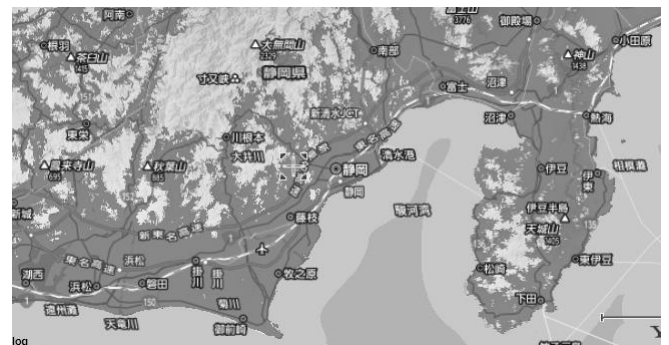


Figure 7: Softbank<sup>TM</sup> communication areas.

The possibility of detecting an unknown spot is described next. For this experiment, we investigated the condition settings for automatic detection of existing spots. However, the tweet amounts during the experiment met conditions in some places other than those of existing spots. For example, spots such as Notsuda Park and Onda River in Tokyo, and Kakegawa Castle Park and Fujikawa Rakuza in Shizuoka were also extracted. Survey results confirmed that, other than spots with correct answers, these are also cherry blossom tourism spots. Further examination of extraction conditions must be done in future studies, but the proposed method can support the discovery of unknown spots.

## 5.2. Best Time Estimation

Results demonstrate that the automatic spot detection method described in 5.1 identified 51 cherry blossom tourist spots. We used the existing method to infer the best times to view cherry blossoms at the detected spots. As an example, we present the famous Meguro River cherry blossoms because the total amount of tweets is large. Using the existing method, we did not experiment when the tweets were few. Therefore, we chose a spot with a large total amount of tweets that is being tested using the existing method. Figures 5–7 present results of the best times to view cherry blossoms on the Meguro River during 2018–2019. Best time estimations, presented as gray, were found. We calculated the one-year moving average for tweet numbers on a daily basis. We calculated the seven-day weighted moving-average on a daily basis. The five-day weighted moving-

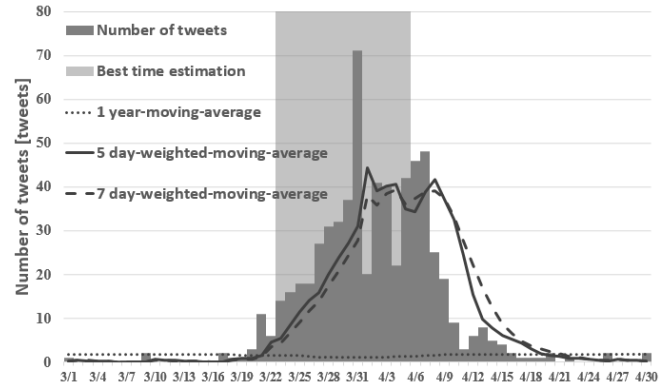


Figure 10: Meguro River in 2019.

Next, Table 5 presents a comparison between the period from the blooming of cherry blossoms in Tokyo to full bloom observed by the Japan Meteorological Agency and the estimated best time to view cherry blossoms along the Meguro River. The period from flowering to full bloom of the Japan Meteorological Agency is indicated by a black arrow. The gray part represents the best time to see it, as estimated using the proposed method. Table 6 presents the recall and precision of each year when the Meguro River cherry blossoms are estimated.

Table 5: Comparison of cherry blossom viewing times

2017	Estimated	2018	Estimated	2019	Estimated
3/10		3/10		3/10	
3/11		3/11		3/11	
3/12		3/12		3/12	
3/13		3/13		3/13	
3/14		3/14		3/14	
3/15		3/15		3/15	
3/16		3/16		3/16	
3/17		3/17		3/17	
3/18		3/18		3/18	
3/19		3/19		3/19	
3/20		3/20		3/20	
3/21		3/21		3/21	
3/22		3/22		3/22	
3/23		3/23		3/23	
3/24		3/24		3/24	
3/25		3/25		3/25	
3/26		3/26		3/26	
3/27		3/27		3/27	
3/28		3/28		3/28	
3/29		3/29		3/29	
3/30		3/30		3/30	
3/31		3/31		3/31	
4/1		4/1		4/1	
4/2		4/2		4/2	
4/3		4/3		4/3	
4/4		4/4		4/4	
4/5		4/5		4/5	
4/6		4/6		4/6	
4/7		4/7		4/7	
4/8		4/8		4/8	
4/9		4/9		4/9	
4/10		4/10		4/10	
4/11		4/11		4/11	
4/12		4/12		4/12	

Table 6: Recall rate and precision rate for cherry blossom viewing

	Recall	Precision
2017	100%	59.1%
2018	33.3%	21.4%
2019	71.4%	31.3%

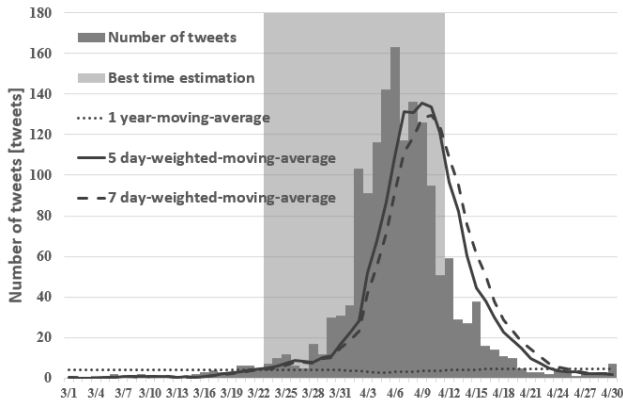


Figure 8: Meguro River in 2017.

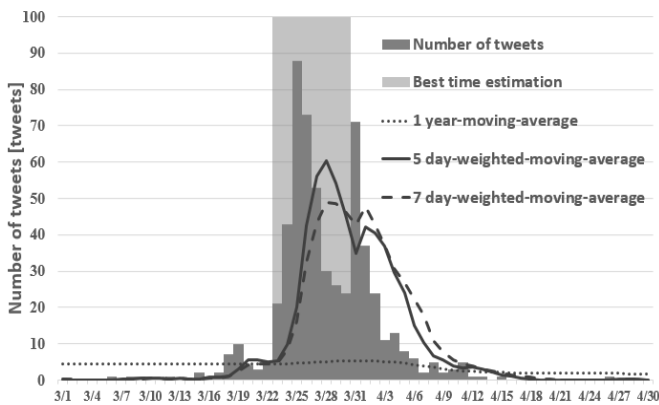


Figure 9: Meguro River in 2018.



From Tables 5 and 6, we obtained good results as best time estimates in 2017 because the period from flowering to full bloom, as indicated by the Japan Meteorological Agency, can be estimated as in full bloom. It is also possible to estimate that it is still in full bloom. In 2019 as well, the recall rate did not reach 100%, but sufficient results were obtained. An estimate can be done without interruption from start to completion of best time estimation. However, in 2018, less than half of the period from flowering to full bloom indicated by the Japan Meteorological Agency was estimated. Probably, estimation from the start to the end of the full bloom was made to some degree. Based on the discussion presented above, estimation results of sufficient cherry blossom viewing times were obtained, except for 2018. The method is somewhat applicable even when used as it is. Of the last three years, two have been successful: in 2018, the recall and precision decreased considerably. Depending on the cause in 2018, it might be difficult to apply the existing method as it is. The next section presents an explanation of causes of the decreased accuracy in 2018.

### 5.2.1. Causes of Reduced Recall in 2018

The decreased recall in 2018 was investigated. Results indicate the weather as a factor reducing the recall rate. Weather information for March 2018 was obtained from the goo weather website [13]. Table 7 shows weather information for March 2018: ● are sunny, ▲ are cloudy, × are rain, and ■ are snow.

Table 7: Weather information for March 2018

Day		11	12	13	14	15	16	17
Temperature (°C)	Max	16.1	15.9	17.5	21.5	22.1	16.9	12.1
	Min	5.5	5.2	5.3	7.2	9	15.7	3.1
Time	9 A.M.	▲	●	●	●	▲	×	●
	12 P.M.	●	●	●	●	▲	×	●
	15 P.M.	●	●	●	●	▲	×	●
Day		18	19	20	21	22	23	24
Temperature (°C)	Max	15.8	18	8.8	3.8	14.5	18.8	16.8
	Min	5.4	8.7	7.7	3.7	3.7	9.4	7.1
Time	9 A.M.	●	▲	×	×	×	▲	●
	12 P.M.	▲	▲	×	■	×	●	▲
	15 P.M.	▲	▲	▲	■	●	●	▲

Table 7 shows sunny weather for March 17, when the meteorological agency indicated the best time for viewing. However, the number of outings probably decreased because the temperature dropped considerably on March 15 and 16 as a result of cloudy and rainy weather. Although the temperature rose on March 18 and 19, the number of tweets did not increase, probably because the weather was cloudy. Then, during March 20–22, bad weather including cloudiness, rain, and snow continued. These effects probably reduced the estimation results for cherry blossom viewing times

considerably. The existing method is intended for best time estimation without consideration of weather information. Therefore, weather strongly influences the estimation result. For the study using the existing method, the best time estimate is that for each prefecture. Therefore, the influence of weather is slight. However, results obtained from this experiment demonstrated that the influence of weather on the tweet becomes strong when one examines one tourist spot unit, as in this case. In other words, one can apply the existing method as it is to estimate the best time to view cherry blossoms in large prefectures. Nevertheless, results demonstrate that noise greatly disturbs estimation results, such as the estimation result in 2018, because the weather effects become stronger when the granularity is fine. Each tourist spot becomes a unit.

## 6. CONCLUSION

This study was conducted to establish a tourist spot detection system requiring no prior knowledge of a tourist area. Earlier research examining this system detected spots that had been featured in magazines and in tweets from the internet. Next, we checked whether the optimal time can be estimated for automatically detected spots. We applied it to a heat map using latitude and longitude information of the geotagged tweet. By applying the conditions there, we detected cherry blossom tourist spots automatically. From experiments in Tokyo and Shizuoka, spot detection rates were 42.2% and 25.0% under condition 1, 60.9% and 40.0% under condition 2, and 79.7% and 62.5% under condition 3. The experiment detected many existing spots, confirming the possibility of automatic detection of tourist sites. Moreover, the discovery of some spots other than the correct answer spot presented the possibility of discovering unknown spots. Next, we estimated the best time to view cherry blossoms in the Meguro River. The recall rates were 100% in 2017, 33.3% in 2018, and 71.4% in 2019. Except for 2018, satisfactory results were obtained. In 2018, its accuracy was reduced greatly because of the influence of weather. The problem persisted. Nevertheless, the system presents opportunities for development because it requires no prior knowledge of cherry blossom sightseeing spots to estimate the best time to view cherry blossoms. In addition, even in spots from which vastly numerous tweets are issued, the weather exerts a strong influence. Therefore, results demonstrated that the existing method using the weighted moving average cannot be applied as it is. The proposed method demonstrated the possibility of automatic detection of cherry blossom tourist spots depending on the setting of conditions. However, the existing method using a weighted moving average was not directly applicable to the detected tourist spots. Therefore, future studies must be conducted to improve the spot detection rate of the automatic detection method of tourist spots and to improve the spot detection rate to assess the spot situation. Furthermore, we expect to examine an optimal time estimation method for specific examination of spot units.

## References

[1] Twitter, <https://twitter.com/>.

- [2] Masaki Endo, Keisuke Mitomi, Keisuke Saeki, Yo Ehara, Masaharu Hirota, Shigeyoshi Ohno, and Hiroshi Ishikawa. 2016. Study of information provided by the best time to see estimation method of phenological observations using tweets. Proceedings of the 12th Annual Conference on Japan Society for Tourism Informatics. Shizuoka, Japan. 47-60 pages. (in Japanese)
- [3] Zen Yoshida and Masayoshi Aritsugi, 2018, Applying a Tendency to be Well Retweeted to False Information Detection, iiWAS '16, 154-159.
- [4] Hiroki Isokawa, Masafumi Toyota, and Yu Kitsuregawa. 2020. Keitaidenwa jinkou toukei data wo mochiita shinki shisetsu ni kakawaru toshi doutai no henka kaiseki, (Analysis of changes in urban dynamics related to new facilities using mobile phone demographic data). Proceedings of the 12th Annual Conference on the 12th Forum on Data Engineering and Information Management (DIEM'20). 75-82 pages. (in Japanese)
- [5] Rocío B. Hubert, Elsa Estevez, Ana Maguitman, and Tomasz Janowski. 2020. Analyzing and Visualizing Government-Citizen Interactions on Twitter to Support Public Policy-making. Digit. Gov.: Res. Pract. 1, 2, Article 15 (April 2020), 20 pages.
- [6] Cheng Guangyao, 2008, Interest Mining and Recommender System for Local Activities, WI-IAT '08, 90-93.
- [7] Shogo Maenaka, Shigeya Morishita, Daichi Nagata, Morihiko Tamai, Keiichi Yasumoto, Toshinobu Fukukura and Keita Sato, 2015, SakuraSensor: a system for realtime cherry-lined roads detection by in-vehicle smartphones, Proceedings of the 15th Annual Conference on the 15th Forum on International Symposium on Wearable Computers (ISWC'15). 345-348pages.
- [8] Twitter Developers, <https://dev.twitter.com/> .
- [9] Meteorological Agency Information on phenology, <https://www.data.jma.go.jp/sakura/data/index.html> .
- [10] Munenori Takahashi, Masaki Endo, Shigeyoshi Ohno, Masaharu Hirota, and Hiroshi Ishikawa. 2019. Study of real-time situation extraction method of tourist spots using SNS data. Proceedings of the 16th Annual Conference of Japan Society for Tourism Information. Tottori, Japan. 33-34 pages. (in Japanese)
- [11] Agricultural Research Institute, Simple Reverse Geocoding Service <https://www.finds.jp/rgeocode/index.html.ja> .
- [12] walker + 1000 views of cherry blossoms nationwide in 2019 <https://hanami.walkerplus.com/> .
- [13] goo weather <https://weather.goo.ne.jp> .

# Radiation Energy of Classical and Nonclassical Vocal Styles

Orie Abe<sup>†</sup>, Rolf Bader<sup>‡</sup>

<sup>†</sup>Information Communication Media Laboratory, Japan

<sup>‡</sup>Institute for Systematic Musicology, University of Hamburg, Germany

<sup>†</sup>orie\_deutschland@hotmail.com

<sup>‡</sup>R\_Bader@t-online.de

## Abstract -

This study deals with the question of radiated energy of the singing voice from the body in general. In addition, it was investigated whether there are differences between classical and nonclassical styles of singing. For the purpose of this research, an experiment was undertaken with seven singers from four various musical genres (classical, musical theater, popular, and Soul singers). The singing voices were measured at 15 parts of the upper body and recorded by a microphone array with 121 microphones while the participants sang five vowels (a – e – i – o – u) at different pitches (90, 120, 180, 250, 380, and 500 Hz), so we also investigated whether the radiation pattern of a singer's voice can be varied according to vowel and pitch. The data gained from the measurement was analyzed from various points of view and displayed in different forms. Our investigation yielded a result that the mouth is usually the strongest sound source in singing, as expected, but strong sound energy radiates from other body regions measured as well. When it comes to differences between classical and nonclassical styles of singing, it was found that the radiation of nonclassical singers is strongly supported by the region of the mouth, whereas the sound energy of the classical singers emitted rather from their whole region measured.

In conclusion, the radiation energy of singing voices depends strongly on the vowels, frequency, pitch and rather on singer's own singing technique than on the musical genre.

**Keywords:** Singing Voice, Sound Radiation, Microphone Array, Minimum Energy Method

## 1 Introduction

The voice has always been used as the most important communication method by humans, and singing has also been evident in everyday life throughout history. Besides the developments in the classical vocal tradition in the 18th century, early forms of the vocal technique of today's popular music, i.e. natural and speech-like singing, began to be practiced intuitively by vocally untrained people as early as the 17th century as well [1]. Each vocal technique developed over time in its particular way, in order to fulfill its intended requirement and associated aesthetics. For example, the instrumental quality of the voice is still most important for classical singing, while nonclassical singing attaches importance to naturalness of the voice [2]. The characteristics of each kind of singing voice are made in the trachea and in the vocal tract, more precisely, involving larynx position, closed phase quotient, subglottal pressure, and vocal tract configuration, and the outcome is

reflected in the frequency, thus, in acoustical characteristics.

However, in spite of such differences of the characteristics among vocal styles, the fact remains, that each singer's voice comes from the body while singing and that the mouth is the main source of its radiation energy. In the process of singing, the mouth is working as a mouth piece, comparable to the mouth piece of a trumpet [3]. But the sound can still be audible, even when it is hummed with closed lips. The energy has to be radiated somehow, otherwise the sound would barely be audible. Another interesting fact is that singers sometimes remark that certain parts of their body tend to vibrate while singing. What about such vibrations? Are these vibrations radiated and what is their strength in the total energy of the singing voice? Or is sound radiation from singers actually restricted to the mouth?

In addition, singers also remark that the strength of such vibrations vary in correlation with the vocal technique, the pitch, and the vowel used. Hence it can be expected that the pattern of sound radiation changes if a singer adopts a different vocal technique or sings a different vowel or tone at a different pitch. But, is that actually so, as most musical instruments have complex patterns of sound radiation, which change with direction, pitches played, and other factors?

Therefore, for a better understanding about the origin of radiation energy of the singing voice, the topic addressed in this study is mainly how strong the radiation energy of the singing voice from the upper body is, in comparison to the mouth radiation, and whether the radiation energy of a singers varies according to vocal technique, pitch, and vowel.

## 2 Previous Research Methods

In the past, the topic whether the vocal sound energy is restricted to the mouth only, has been researched repeatedly (e.g. Kirikae et al. [4], Fant et al. [5], Sundberg [6], Pawlowski et al. [7], Sakakura et al. [8], and Takada [9]). These studies found out that the sound vibration depends on vowels because of the mouth aperture size and on pitch. Furthermore, stronger vibration was also observed at loud vocal sound.

The previous studies on this topic have mostly been executed using an accelerometer, with which the energy of each desired part of the body has been measured. In order to yield an accurate result from each desired part of the body, a measurement using an accelerometer is certainly a suitable research method. But our investigation was executed by means of a microphone array with 121 microphones, the so-called Acoustic Camera, because using a microphone array has two

advantages which would be ideal for the purpose of this research [10]:

- It enables us to see the radiation field of the radiating source.
- The recorded data can be used to back-propagate the sound field to the surface of the radiating source.

The method which uses a microphone array is usually used in musical instrumental acoustics nowadays. However, in vocal acoustics, the use of microphone grid seems to be standard and the research method using an acoustic camera is still relatively new, particularly in the size with 121 microphones.

### 3 Aim, Method and Research Materials

#### 3.1 Aim

Mainly, this study attempted to find an answer to the following questions in two research ways:

##### Research 1

- Is it true that the main sound energy of the singing voice only radiates from the mouth? Or are there other body parts involved in the sound radiation of the singing voice?
- If so, are there recognizable differences in vocal music genres /singing techniques and vowels?

##### Research 2

- How strong is the sum of the radiated energy from all the measured parts of the body in relation to the mouth radiation?

Additionally, another analysis was executed, because the radiation energy originated from the corners of the mouth has presumably a strong influence on the total energy, due to the location of these parts close to the mouth. Therefore, the question is:

- What about the radiation energy excluding the corners of the mouth?

#### 3.2 Method

The microphone array, which was employed in our investigation, consists of 121 microphones (11 x 11) and the array spacing is a regular grid with a grid constant of 3.9 cm. This construction enables a symmetric visualization of radiating field later. The microphones record simultaneously with a sampling frequency of 48 kHz, thus the whole human hearing range up to 20 kHz is covered. This is a very important factor for measurements of musical instruments including singing voice, because musical instruments often radiate high frequency and initial transients are often the most important part of the sound [10].

After the recording, the data obtained by the array were used to back-propagate the sound field to the radiating source

surface by means of a near-field method, the so-called Minimum Energy Method, for reconstructing a sound pressure field at a radiating surface of musical instruments including the human body for voice research. It samples the source plane by as many the so-called equivalent sources as the microphones present in the Acoustic Camera. This method enables a reconstruction of sound pressure fields and a visualization of an overall radiation directivity of a vibrating geometry.

For the analysis, a code written in Mathematica was applied to all the data and the vibrations were analyzed on a total of 15 parts of the upper body (mouth / chin / throat / left and right clavicles / sternum / nose / nasal bone / left and right corners of the mouth / left and right cheeks / forehead / left and right lower eyelids). This setting made it possible to show energy values of the voice radiation from the singer's upper body, including the phase angles and all statistical procedures are based on those visualized results. The data are reconstructed and visualized on a photo of the human upper body from the head to the chest.

Furthermore, the strongest radiated area was color-marked: the radiation is adjusted up to -6 dB and the intensity of the radiation energy is visualized by colors (the brighter the color, the stronger the radiated energy). Usually the radiation from the mouth is the strongest, so that the radiated energies of all single frequencies were therefore normalized to 0 dB at the mouth, but there were some figures where energy value at the mouth is shown in -1 dB. This means that the revealed energy value is not quite 0 dB, so somewhat stronger than 0 dB.

Both, this Acoustic Camera and the Minimum Energy Method, were developed at the Institute for Systematic Musicology, University of Hamburg, and all information about these recording and analysis techniques can be found in the publications of the developer Rolf Bader (Detailed description about the Acoustic Camera and the Minimum Energy Method as well as exemplary measurements can be found e.g. in [10] [11] [12] [13] [14]).

#### 3.3 Research Materials

Seven trained singers — three classical, two musical theater, and two popular singers — participated in this study and one popular singer's voice (subject: VS) was recorded twice by means of popular singing technique and by means of Soul singing technique:

1. CH (classical, female, alto)
2. SE1 (classical, female, soprano)
3. JR (classical, male, bass)
4. SS (musical theater, female, mezzo soprano)
5. TF (musical theater, male, tenor)
6. SE2 (Pop, female, mezzo soprano)
7. VS (Pop, female, mezzo soprano)
8. VS (Soul, female, mezzo soprano)

For the study, after a short warming-up vocal exercise, each subject sang five vowels a/e/i/o/u at the following notes;

- 90 Hz (bass), analyzed up to 1,3 kHz (15 partials)
- 120 Hz (males), analyzed up to 2,5 kHz (21 partials)
- 180 Hz (alto/mezzo), analyzed up to 3 kHz (17 partials)
- 250 Hz (all subjects), analyzed up to 4 kHz (16 partials)
- 380 Hz (tenor), analyzed up to 4,5 kHz (12 partials)
- 500 Hz (females), analyzed up to 5 kHz (10 partials)

Each vowel was recorded separately and max. 2 seconds of phonation. In order to get good data, the vowels were repeatedly measured two or three times in the order of a-e-i-o-u, respectively. The measurement was executed in an anechoic chamber. For the recording, the microphone array was attached to the front of a stand and adjusted for the height of the singer, so that their mouth is positioned in front of the center microphone (No. 61). As already mentioned, the Minimum Energy Method is a near-field method, therefore the center microphone was placed 3 cm in front of the mouth.

## 4 Results

As already mentioned in Subsection 3.1, the data gained from the measurement was analyzed from various points of view and displayed in different forms, so that the research findings will be shown in two ways.

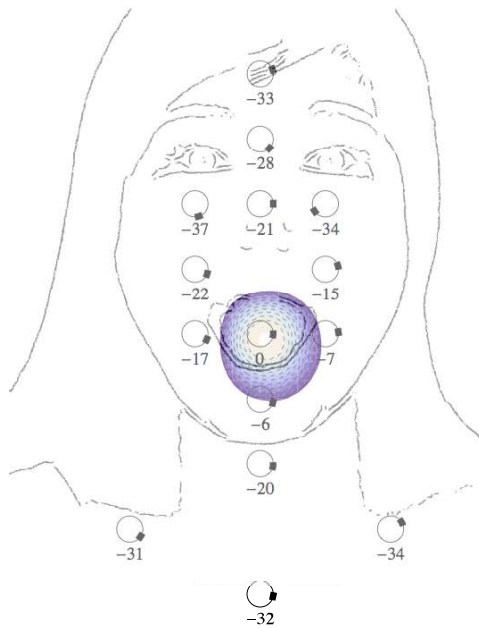


Figure 1: Radiation of a classical soprano (subject: SE1) at the fundamental for the vowel /a/ at 250 Hz.

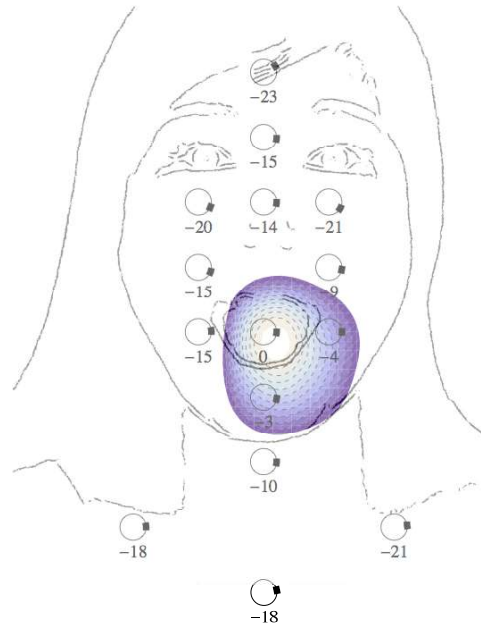


Figure 2: Radiation of a classical soprano (subject: SE1) at the fundamental for the vowel /a/ at 500 Hz.

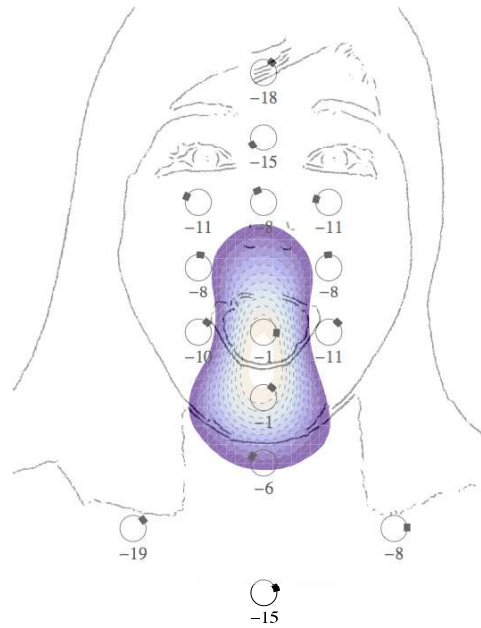


Figure 3: Radiation of a female Pop singer (subject: SE2) at the fundamental for the vowel /e/ at 250 Hz.

## 4.1 Source of the Radiation Energy in Singing (Results of the Research 1)

At the fundamentals, it seems for all the subjects that the strongest radiation energy of their singing voice (color-coded area) came uniformly from the mouth, as shown in Figure 1 as an example. It was already expected that the mouth is generally the strongest radiation source of the singing voice. In addition, the corners of the mouth often vibrated strongest of all the measured parts except the mouth due to their spatial closeness to the mouth, in fact independent of vowel. In general, it was revealed that the closer to the mouth the measured part of the body, the stronger the radiation energy.

When singing at higher pitches, the radiation energy increased in total in comparison to that revealed from singing at lower pitches, so that strong radiation energy came from a large area of the lower part of the face (see Figure 2). Therefore, by means of visualizing the strongest radiating area up to -6 dB that was marked in color, it is clearly noticeable that the area of strong radiated energy shown at the fundamental became large with increasing pitch that was sung.

The finding that for singing at high pitch, the radiation of the singing voices became stronger at high frequencies, is presumably caused by a high lung volume and high-speed opening and closing of the vocal folds at the pitch, so that strong sound pressure will be produced by these factors. This is just what Sundberg asserted, although the body regions concerned are not confirmed by his recognition: very strong sound pressures in the vocal tract and in the trachea generate the phonatory vibrations in the skull, neck, and chest regions [15].

Because of the higher sound pressure caused by singing at high pitch, it was observed that the radiation energy from some parts of the body can be stronger than that from the mouth at higher partials, as shown in Figure 3. Such radiation energy was observed by shifting the energy source or emergence of multiple energy sources that were often found in the case of 500 Hz, especially. This can be a temporary, but also as a continuing phenomenon.

The radiation patterns observed in our study revealed that the sound radiation changes depending on the frequency, and that the energy from the region of the body outside of the mouth increases strongly up to about 3 kHz, as displayed in Figures 4 and 5. The chin and the corners of the mouth were the parts showing the strongest radiation of energy of the 14 measured parts of the body (except for the mouth radiation), and the radiation from the chin often remained almost unchanged in the whole frequency range analyzed. However, it was found that the radiation energy of the parts of the body that are located far away from the mouth, increases dramatically and fluctuates strongly compared to the parts which are in the place near the mouth.

The strong radiation energy from the measured parts of the body at high frequencies is presumably related to the fact that the partials at high frequencies radiate forwards, whereas these at low frequencies emit almost equally in all directions, as reported by Marshall and Meyer [16].

However, when it comes to the dimension of energy increase, there were large differences among the vowels and

individuals. While the progress of the vowels /a/e/i/ were a smooth increase, that of the vowels /o/ and /u/ showed a strong zigzag-like course (see Figures 4 and 5. The energy from the mouth is shown at zero on the x-axis (dark blue line).). This fluctuation was visible at all the parts of the body in many cases. Presumably, there were complicated air motions in the vocal tract due to the narrowing of the mouth opening and configuration changes of the vocal tract for these vowels.

As mentioned above, the dimension of the energy increase was also different in each individual case. Some subjects showed a strong energy value from the beginning, i.e., from the fundamental frequency, so that their energy only slightly increased in comparison to the rest. For example, for classical singers, their energy levels gained from all the parts of the body were relatively close to each other for all the vowels, so that the difference of these energy levels was smaller among the measured parts in comparison to the nonclassical singers (see Figures 4 and 5). The energy from the mouth is shown at zero on the x-axis (dark blue line).). Therefore, for the classical singers, there was no large increase of the energy from the region of the body outside of the mouth in the frequency range analyzed, in comparison to the energy of the nonclassical singers. This implies that the energy of classical singers radiates rather from their whole upper body while singing.

In regards to the difference in musical genre, the radiation energy of the nonclassical singing techniques for the vowels /o/ and /u/ strongly originates from the mouth area, i.e., from the corners of the mouth and the chin. This is clearly noticeable when looking at the progress of the radiated energy, because there was a large difference in the energy level between these parts of the body and the rest of the parts measured, as shown in Figure 5 (Top three lines are the results from the corners of the mouth and the chin). Presumably, this is due to different mouth/lip opening: classical singers usually hold the shape of their mouth/jaw opening as constant as possible for all vowels in order to keep certain vocal loudness and beauty of the voice, independently of the changing pronunciation of vowel sound. This difference in musical genre was observed distinctly at low pitch than at high pitch that is sung, likely due to a wider mouth opening at higher pitch.

## 4.2 Total Radiation Energy from Measured Parts of the Body in Comparison to the Radiation Energy from the Mouth (Results of the Research 2)

When it comes to the sum of the radiated energy from all the measured parts of the body in relation to the mouth radiation, all the subjects showed that the total energy of their singing voice can exceed the energy from the mouth (see the result above in Figure 6. The result was compared to the mouth radiation, thus, the line above zero point means that the total radiation energy is stronger than the radiation energy from the mouth.). But the energy level strongly depended on the frequency analyzed, the vowel and the pitch that was sung, just like the research findings shown in Subsection 4.1.

Comparing the radiated energy from the mouth with the total energy from all the parts excluding the energy from the



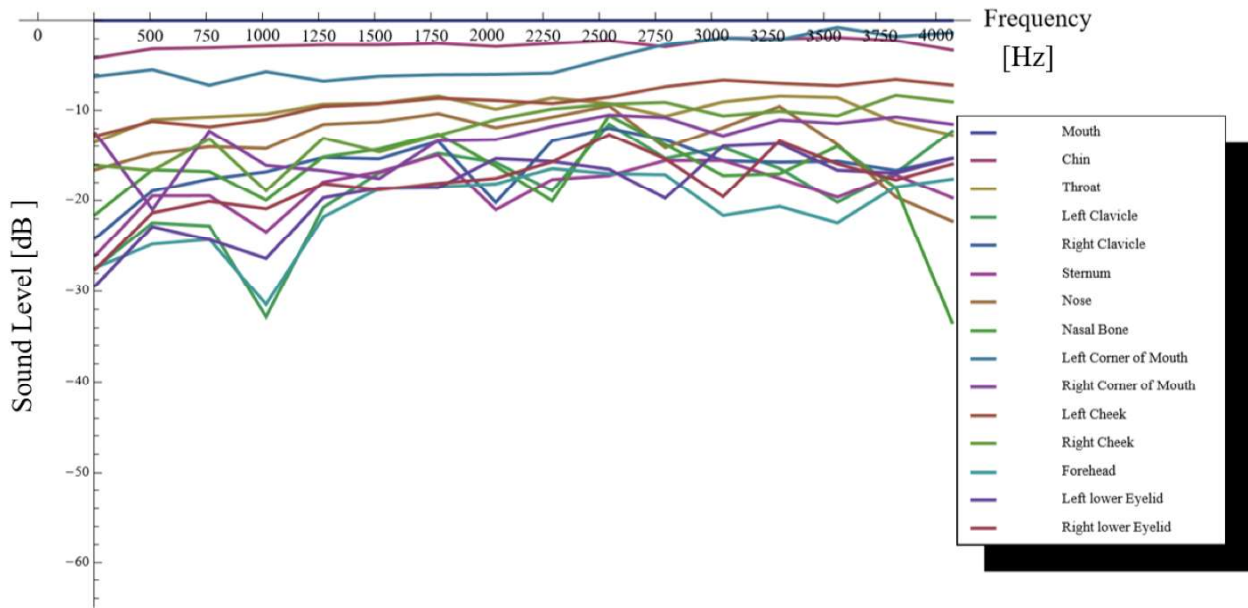


Figure 4: Sound levels of 14 measured parts of the body sung by a classical soprano singer (subject: SE1) relating to the radiation from the mouth (0 dB) at 250 Hz for the vowel /a/.

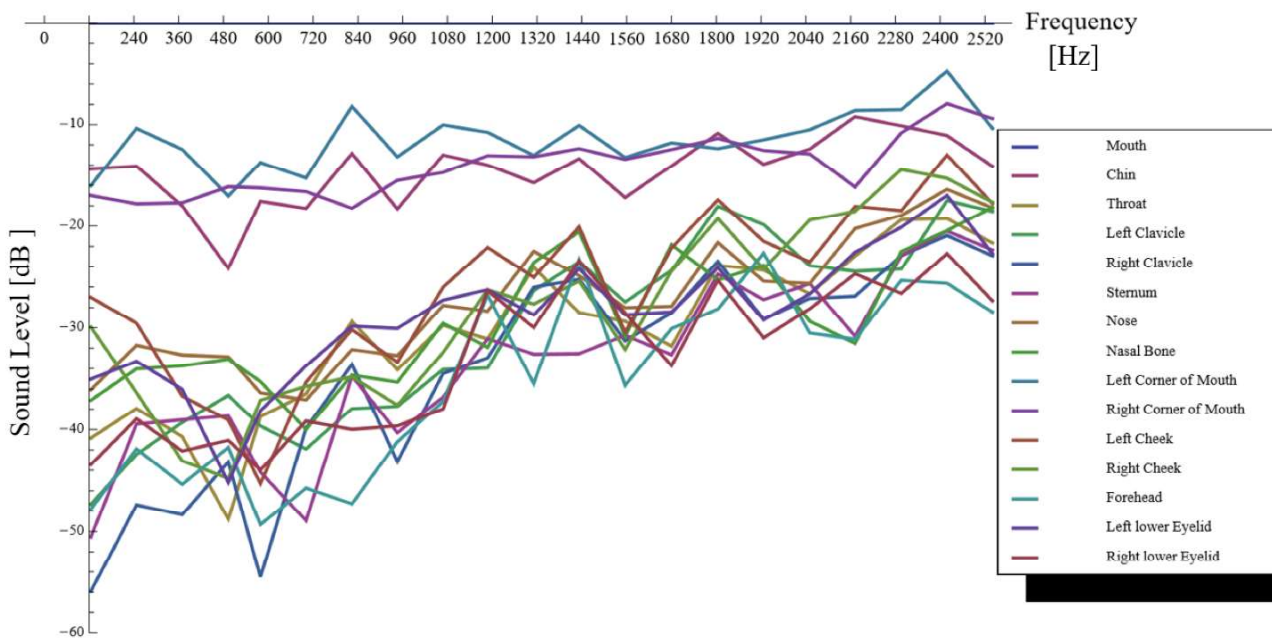


Figure 5: Sound levels of 14 measured parts of the body sung by a male musical theater singer (subject: TF) relating to the radiation from the mouth (0 dB) at 120 Hz for the vowel /u/.

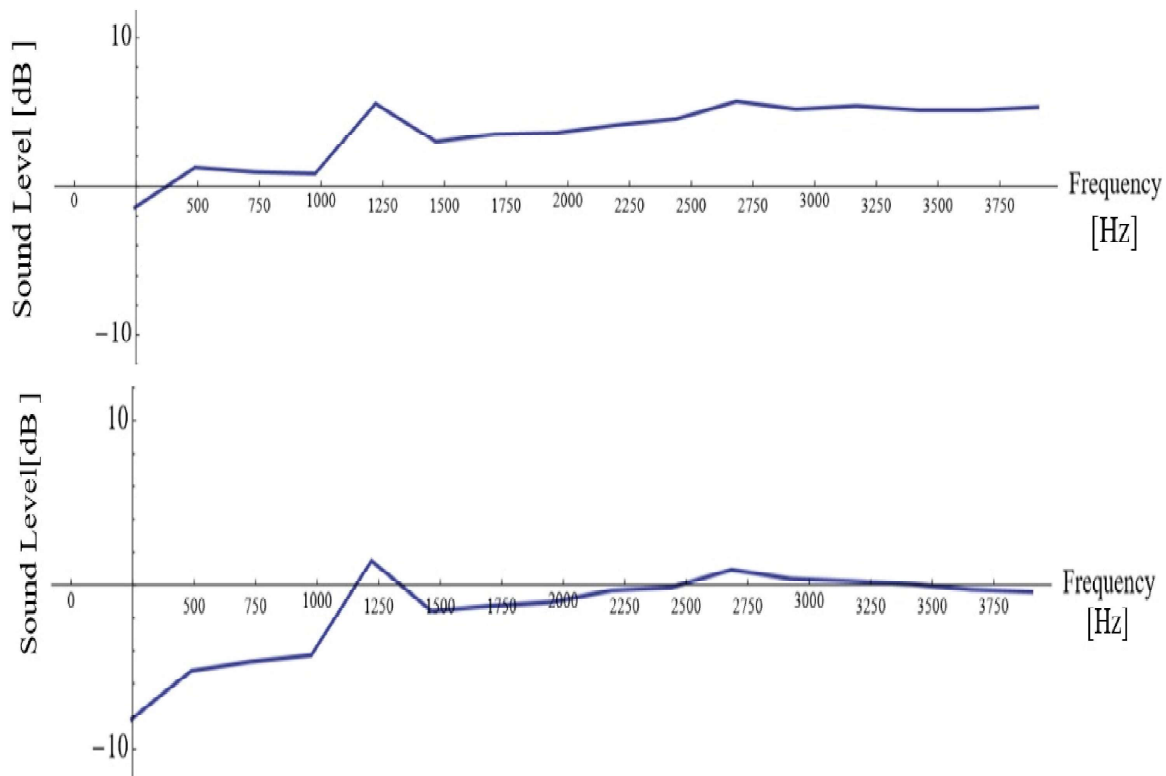


Figure 6: Total radiation energy from 14 measured parts of the body (top) and the total energy excluding the energy from the corners of the mouth (bottom) compared to the mouth radiation sung by a female Pop singer (subject: SE2) for the vowel /a/ at 250 Hz.

corners of the mouth (i.e., energy from the following parts is included in this total energy: chin / throat / left and right clavicles / sternum / nose / nasal bone / left and right cheeks / forehead / left and right lower eyelids) made it clear that the energy from the corners of the mouth has a strong influence on the total energy of the singing voice (see the result below in Figure 6).

But, in spite of the absence of the radiation energy from the corners of the mouth, it was shown that the total energy can also exceed the energy from the mouth in general. This emerged either as a temporary occurrence or as a continuing phenomenon at higher frequencies.

## 5 Conclusion

This study revealed clearly that the mouth is usually the strongest sound source in singing, as already expected. However, sufficiently loud sound pressures from other parts of the body, at least from the fifteen parts measured, were also observed in this investigation. In most cases, an increase of the energy from these parts was shown up to about 3 kHz for all the vowels, even though the energy level showed a rather changeable process at all the pitches for the vowels /o/ and

/u/. This increase depended on the pitch that was sung.

Because of this increase, in some cases, a few of the parts of the body revealed even higher energy values than the mouth radiation at high frequencies. In addition, we found that the total energy from the 14 parts even surpasses the radiation energy values from the mouth. This depended strongly on frequency, but this fact is a remarkable finding for a better understanding about the origin of radiation energy of the singing voice.

Furthermore, it was clarified that the energy from the corners of the mouth has strong influence on the total radiated sound energy level by comparing the energy from the mouth with the total energy from the other parts excluding the energy from the corners of the mouth. This influence was expected, due to the spatial closeness to the mouth.

In other aspects, it was found that total radiation energy of the singing voice is strongly supported by the enhanced energy from the body regions that are located far from the mouth, because a dramatic increase in energy was found there, and because the energy values from the body regions of the chin, the corners of the mouth, and the cheeks did not change as much. However, for the vowels /o/ and /u/, the radia-



tion energy of nonclassical singers is strongly supported by the corners of the mouth and the chin. For the nonclassical singers, the difference in the radiation energy level between the region of the mouth (i.e. the corners of the mouth and the chin) and the rest of measured parts of the body was much larger than for classical singers. This finding indicated that classical singers use rather their whole body as a musical instrument, therefore the radiation energy of their singing voice emits steadily from there, in fact, independent of the vowel, for example.

In conclusion, the results of this study revealed that the radiation energy of the singing voice emits not only from the mouth, but also from other regions of the body. However, the radiation energy of the singing voice depends strongly on the vowels, frequency, pitch, and person who sings. This means that the radiation of the singing voice depends rather on the singing technique of each singer than on the musical genre.

We hope that our research findings can give singers and vocal teachers of various musical genres a helpful information for their vocal training and teaching.

## REFERENCES

- [1] Sadie, Stanley E.: The new Grove dictionary of music and musicians, Groves Dictionaries Inc, 2nd edition (2001).
- [2] Schutte, Harm K. and Miller, Donald G.: Belting and pop, nonclassical approaches to the female middle voice: some preliminary considerations, *Journal of Voice*, Elsevier, Vol.7, No.2, pp.142–150 (1993).
- [3] Titze, Ingo R.: The human instrument, *Scientific American*, Nature Publishing Group, Vol.298, No.1, pp.94–101 (2008).
- [4] Kirikae, J., Sato, T., Oshima, H. and Nomoto, K.: Vibration of the body during phonation of the vowels, *Revue de Laryngologie, Otologie, Rhinologie*, Vol.85, pp.317–345 (1964).
- [5] Fant, G., Nord, L. and Branderud, P.: A note on the vocal tract wall impedance, *Speech Transmission Laboratory Quarterly progress and status report*, Vol.4, pp.13–27 (1976).
- [6] Sundberg, J.: Chest wall vibrations in singers, *Journal of Speech, Language, and Hearing Research*, Vol.26, No.3, pp.329–340 (1983).
- [7] Pawlowski, Z., Pawloczyk, R. and Kraska, Z.: Epiphysis vibrations of singers studied by holographic interferometry, *Proc. SMAC-83*, Royal Swedish Academy of Music, Vol.28, pp.37–60 (1983).
- [8] Sakakura, A. and Takahashi, H.: Body wall vibration in trained and untrained voices, *Vocal Physiology: Voice Production, Mechanisms and Functions*, (Fujimura, O., ed.), Raven Press, New York, pp.391–401 (1988).
- [9] Takada, O.: Radiation Pattern of Classical Singing Style, *Systematic Musicology: Empirical and Theoretical Studies*, (Schneider, A., v. Ruschkowski, A., ed.), *Hamburger Jahrbuch für Musikwissenschaft*, Peter Lang, Frankfurt/M, Vol.28, pp.163–173 (2011).
- [10] Bader, R., Münster, M., Richter, J., and Timm, H.: Microphone Array Measurements of Drums and Flutes, *Musical Acoustics, Neurocognition and Psychology of Music*, (Bader, R., ed.), *Hamburger Jahrbuch für Musikwissenschaft*, Peter Lang, Frankfurt/M, Vol.25, pp.15–55 (2009).
- [11] Bader, R.: Reconstruction of radiating sound fields using minimum energy method, *The Journal of the Acoustical Society of America*, Vol.127, No.1, pp.300–308 (2010).
- [12] Bader, R.: Microphone array, *Springer Handbook of Acoustics*, Springer, pp.1179–1207 (2014).
- [13] Bader, R.: Radiation characteristics of multiple and single sound hole vihuelas and a classical guitar, *The Journal of the Acoustical Society of America*, Vol.131, No.1, pp.819–828 (2012).
- [14] Abe, O.: Sound Radiation of Singing Voices, <http://ediss.sub.uni-hamburg.de/volltexte/2019/9954>, Dissertation, Universität Hamburg, (2019).
- [15] Sundberg, J.: Phonatory vibrations in singers: A critical review, *STL-QPSR*, Vol.32, No.1, pp.37–51 (1991).
- [16] Marshall, AH. and Meyer, J.: The directivity and auditory impressions of singers, *Acta Acustica united with Acustica*, S. Hirzel Verlag, Stuttgart, Vol.58, No.3, pp.130–140 (1985).



Session 3:  
Systems and Services  
( Chair: Yoshia Saito )



# Process Improvement Method for Progress Meeting

Akihiro Hayashi<sup>†</sup>

<sup>†</sup>Department of Information Design, Shizuoka Institute of Science and Technology, Japan  
pixysbrain@gmail.com

**Abstract** - While work style reforms have been discussed in different fields, an area that demands improvement is the conference process of white collar workers. The meeting culture in each company is one that has been cultivated for several years, and hence, was difficult to improve rapidly. In this research, some causes of the problem were identified by analyzing the meetings held in several companies. In order to address these causes, we proposed a new method of integrating best practices and introducing free IT tools and AI applications that can be used at a low cost by PCs and smartphones. When the proposed method was applied to actual progress meetings, the effects of improvement could be confirmed in terms of labor saving, time reduction, and paperwork reduction.

**Keywords:** Meeting Process, Smart Phone, Free AI Applications

## 1 Introduction

According to the data of the Ministry of Economy Trade and Industry in 2019, approximately 60% of the workers in Japan are so-called white collar workers, and most of them do routine intensive work in offices. There are reports that about 37% of their working hours are used for meetings[1]. Therefore, if conference operations, which are held by allocating multiple working hours, are not controlled efficiently, the productivity of the entire company could be affected.

In the manufacturing industry, project efforts are classified into three categories: management, regular meetings, and manufacturing. We consider that the proportion of the management effort at 15% is reasonable in total effort. If 37% of the effort is spent on regular meetings, the time devoted to manufacturing is only 48%. Management and meetings both incur indirect costs, without generating any income. Thus, reducing meeting effort is viewed as a crucial issue.

There are several ways to reduce the meeting time. The first method is to avoid meetings that continue without any sense of purpose. Even in productive meetings, reduction of the types and frequency is a possibility.

Meetings held in companies are roughly divided into formal and informal meetings. Formal meetings are often held at a regular frequency, such as annually, quarterly, monthly, weekly, or daily. They also operate according to a predetermined process. Such official meetings are called routine work. In contrast, informal meetings do not require any special meeting procedure and generally, no conference minutes are prepared.

In this study, the discussion is focused on formal meetings. A conference held at a company is defined as a process. By improving the conference process, we propose to enhance the productivity of the conference, reduce the time and energy requirement, and operate sustainably in a paperless manner.

In an earlier study, Tomobe et al[2]. proposed a disc cash media as a mechanism for efficiently sharing the conference structure as well as the discussion. We have constructed a system that enables viewing in an interactive manner as well as a method of making it possible.

Hiramitsu et al[3]. focused on improving communication at conferences. An experiment was conducted to examine the participants' judgment and thoughts regarding the speaker's opinions, and to examine the degree of speech, predominance, and efficiency in a conference that also included chat. It was revealed that the participation level was not proportional to the number of utterances. Expectations were recognized as a place to reduce resistance to chatting, speak on behalf of someone, and respond to chats.

Sengoku et al[4]. focused on interactions such as conferences and meetings, which were cost factors, and measured and evaluated their transaction effects. Here, the different performance indicators and behavior indicators for the organizers and participants of the conferences and meetings were investigated. Subsequently, multiple significant correlations and causal relationships were extracted between them by statistical verification and the results were interpreted by structural equation modeling.

Yamamoto et al[5] studied a brainstorming-type workstation that is used at the conceptual stage as a means of communication in the project. Meetings for deciding significant matters and progress meetings were considered, and the ideal minutes for maintaining the quality of the project were discussed.

However, no particular discussion was initiated that suggested a best practice for the meeting process and applied free IT and AI techniques to improve the meeting process.

In this paper, we have defined the issues of meeting productivity in section 2 and clarified the root cause of the issues using the case analysis method. Section 3 proposes a method for solving the issues identified in Section 2. In Section 4, the proposed method is applied to some conferences held in actual universities to evaluate its effectiveness. We will discuss this in Section 5 and describe the conclusion in Section 6.

## 2 Issues of Meeting Process and Cause Analysis

### 2.1 Issue to be solved

The main purpose of any meeting must be the reporting of necessary items and decision making. In addition, meetings may be conducted for communication purposes, within the department. Since the manner of holding meetings is a culture that each company has cultivated over several years, adverse effects may occur if the frequency and/or duration of the meetings are effortlessly shortened.

Firstly, a productive way to manage the conference while reducing the variety and frequency of the meetings must be investigated. The issue is to establish a meeting process that conducts the meetings effectively and efficiently.

### 2.2 Analysis of Factors that Inhibit Productive Meetings

In this study, to identify the causes that prevent productive meetings, a case analysis of the meetings held by a few organizations was conducted. An outline of the methods followed by each organization is provided as follows.

**Case1** NTT affiliates: There are a significant number of meetings and numerous procedures. The number of meetings increases every time a problem occurs.

**Case2** Foreign capital computer company: A sharp distinction is observed between official and informal meetings. Lotus Notes are distributed to all employees and conference materials are shared. Meetings are strictly paperless.

**Case3** Manufacturer of precision equipment for automobiles: Under the management of the founding family, only the chairman and president have authority and all other employees are yes men.

**Case4** Private science university: Strict lifetime employment and no restructuring. All attendees have the authority to speak at these meetings.

From consultations and documents such as session minutes from the aforementioned organizations, a fish-bone chart was created as shown in Fig.

From the analysis, three factors were recognized as the obstacles to conducting productive meetings.

#### 1. Rework Occurrence (Factor 1)

At formal meetings, work products and meeting minutes are created during and after the meeting, respectively. However, work products created in this way are seldom approved immediately. After the meeting, perception gaps are identified.

#### 2. Difference in Perception of Meetings (Factor 2)

Meetings are held for decision making. Accordingly, it is vital to understand the background of the meeting issue. However, contents of the material will be explained at the conference.

#### 3. Complicated Meeting Procedure (Factor 3)

There are several formal procedures to be followed. The minutes of the meeting are rewritten with numerous comments and then approved and distributed.

## 3 Process Improvement of Meeting Process

In Section 3, we propose a continuous improvement method for the meeting process to solve the issues pointed out in Section 2. Various types of meetings are held at the organizations like, management meeting, staff meeting, section meeting, and project progress meeting (hereinafter abbreviated as progress meeting). The process of a meeting depends on its character. It is difficult to conclusively identify the best practice for each conference process.

This research primarily focused on progress meetings. There is only one management meeting for the entire company, but there are several weekly progress meetings owing to multiple projects. If a method to manage the progress meeting effectively and efficiently is established, considerable benefit may be expected.

Progress meetings are held according to the project plan. The meeting is held on a weekly basis and the progress of the task assigned and problems of the previous week are shared among project members.

### 3.1 Basic Policy

In order to continuously improve the meeting process, the following two basic policies are adopted in this research.

#### 1. Continuous improvement of the meeting process

We identify the rework of the meeting process shown in Factor 1 and aim to continuously improve the meeting process using PDCA to eliminate the rework. This leads to an elimination of factors 1 and 2.

#### 2. Utilize free IT and AI applications

We will eliminate the effort by positively utilizing AI applications installed on PCs and smartphones. This will lead to an elimination of factors 1 and 3.

### 3.2 Proposal for Progress Meeting Process

In Section 4, we propose the meeting process improvement method to address the issues and factors mentioned thus far. By repeating the PDCA management cycle as defined below, we aim to improve productivity by eliminating any rework of meetings.

- PLAN Creating a meeting plan and rules

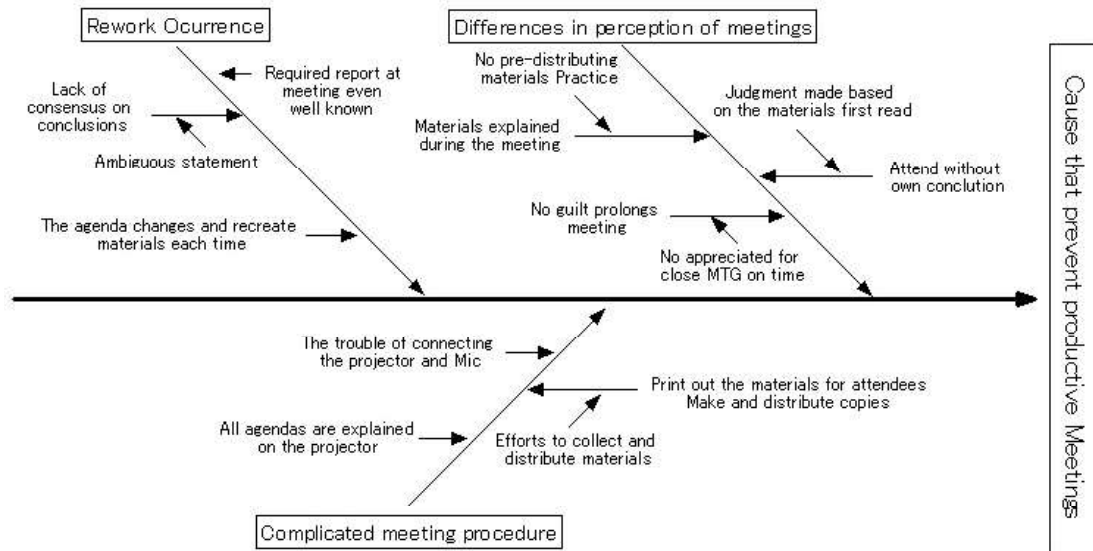


Figure 1: Factor that prevent productive Meetings

- DO Enforcement of basic actions before and during the meeting
- CHECK Evaluation and Measurement
- ACT Adding practice to the meeting process

Currently, more than 80% of desktop PCs, use Microsoft Windows. Recently, an increase has been observed in the use of Apple's Mackintosh for business purposes. Irrespective of the OS used, Microsoft Office 365 is often installed and common applications like Word, Excel, Powerpoint, Outlook are used. If your organization has Office 365 with a business license, 1TB disk capacity on the cloud can be used with OneDrive.

The following proposal will be explained based on the assumption that Microsoft Office 365 has been in all PCs, and the applications and hard disk capacity can be used.

### 3.2.1 PLAN Creating a meeting plan and rules

The outcomes of the weekly progress meetings are reported to the management meeting through mid-level managers. The following is an explanation based on the assumption that a management meeting will be held every Monday morning.

#### 1. Meeting Plan

Holding a conference is also considered a management procedure. A meeting is planned and conducted accordingly. 5W1H of the progress meeting is described in the meeting plan. Since the management meeting is on Monday, the progress meeting will be held every Friday between 1-3 pm. The conference room will be reserved for up to 6 months, along with equipments

such as projectors and IC recorders. In principle, every weekly progress management has the same items on agenda; thus, an agenda is set and placed in the date folder. The agenda has Report Items and Discussion Items. Progress data such as the EVM value is entered in advance. In the Discussion Items, project-related problems are described and necessary decisions are made.

#### 2. Notice of Progress Meeting

Notice of the progress meeting is delivered via Outlook. In previous studies, it has been reported that organizations that utilize Outlook on a daily basis, have achieved an average time saving of 100 h per year by using Outlook shortcuts and procedures[4]. Since notification of the Progress Meeting generally has similar content, an Outlook template is used and modified as required.

#### 3. Configuration Plan for Meeting Materials

The Office365 Business License allows OneDrive to allocate 1TB of disk capacity for each user. The space for the progress meeting is secured in OneDrive, and the date folders for 6 months are created in advance. The meeting agenda and materials for each meeting are placed, and the OneDrive folder is shared to each employee's PC. Each participant uses the original on the cloud as it is. Printing or copying of electronic files is prohibited.



### 3.2.2 DO Enforcement of basic actions before and during the meeting

1. Update EVM  
EVM (Earned Value Management) is used to track the progress. EVM is designed to measure value based on time or money. In this research, time is defined as a measured value. The weekly work is measured in hours on Friday morning, and updated WBS's earned value (EV) and actual cost(AC). When the EVM is updated for all the project tasks, the EVM parameters are automatically updated.
2. Update Agenda  
For the reporting items that are not needed to discuss, the progress information including the EVM value are entered. Discussion Issues like risks, concerns regarding dependencies, etc. are described in the discussion items. If you notice this issue during project management, open the progress meeting agenda on the cloud and register it each time by event driven.
3. Prior review of meeting materials  
On Friday morning, the attendees will read the agenda and meeting materials in advance. During the progress meeting, the materials in the date folder must be understood in principle. Explanation of the materials during the meeting is prohibited.
4. Operations during meetings  
When the progress meeting was held on Friday afternoon, the chair opened the OneDrive date folder and assigned the registered document number. Then, according to the agenda, the project members reported the status of the project and any problems/risks/issues. In principle, issues were supposed to be registered in advance and thus, the items were briefly reported and evaluated.
5. Updating an independent management ledger  
Progress data and questions asked are written in the agenda directly. However, progress meetings are held weekly and electronic documents are stored in the OneDrive date folder. Certain issues and agenda continue onto the subsequent weeks and considerable effort is spent to search previous folders, thus leading to immense rework. Therefore, a ledger independent of the progress meeting is created for issues, risks, and dependency management. This ledger is not created for each local project, but for all members who hold the progress meeting.
6. Update ledger by voice input  
When updating work products such as the issue management ledger during the progress meeting, the documents are projected onto the screen and the participants input voices on their smart phones. If the PC and smart-phone are in the same Wifi environment at a specific

location, a remote mouse can be used to allow multiple participants to input their voices simultaneously. This function is used to update the issue management ledger. Then, the time and effort required for keyboard input are saved.

7. Creating minutes by automatic transcription  
The progress meeting is limited to a maximum of 2 h. All utterances under discussion are recorded on the IC recorder and transcribed using AI after the conference. In this proposal, the meeting result is recorded on the progress meeting agenda, and no new meeting minutes are created. However, a quick record is created as a backup to avoid ambiguous parts. The method is to set the stereo mixer on Windows and set "Record" to input audio through the PC. In newer Windows, in the absence of a stereo mixer, the free software VB-CABLE Virtual Audio Device can be used. For Macintosh, use Sound Flower.
8. Report using issue management ledger  
After holding the progress meeting at 1-3 pm on Friday, the chairman will report the result to the senior manager prior to 5 pm on Friday. Regarding the report to the upper level manager, the problem management ledger was used to report a critical problem in about 20 min. Therefore, if suggested by a senior manager, homework items were assigned to the members on Friday. In addition, the senior manager also provides the authority to view the progress meeting folder so that the progress meeting documents can be viewed at any time.

### 3.2.3 CHECK Evaluation and Measurement

1. Measuring meeting time  
In the proposed method, it was not necessary to explain the material during the conference as in the past. Instead, all participants read the documents prior to the meetings. As a result, the meeting time itself is not reduced, but the burden of advance preparation increases. It is considered to incur the same cost when the creator of the material spends 1 h explaining to the 10 participants during the meeting, and when each of the 10 members spends 1 h to read the material prior to the meeting. Based on that premise, we measured the time required for conferences and measured the increase and decrease in time due to the adoption of the proposed method.
2. Measuring rework time  
The decisions taken in the progress meeting are recorded in minutes and the assigned task is managed by using the action list. However, after a week, it becomes difficult to understand the contents of the minutes alone. The original date folder may be opened, searched, and read again. In this study, we define this time as rework. This rework time was measured.

### 3.2.4 ACT Add practice to the meeting process

#### 1. Refine the results of measured reword time

If the conference time exceeds the previous one, it implies that the plenary conference process is not convenient. On analyzing the breakdown of the meeting time and rework time measured by HECK, identifying the bottlenecks helps in improving any faulty practices.

## 4 Results and Evaluation of Issue solving

Through a search seminar in the university laboratory, the proposed method was applied and evaluated. University seminars are similar to progress meetings as they meet once a week and report on progress over a specific project period. The effectiveness of the proposed method was evaluated by faculty and students of our laboratory by assigning roles and introducing a process of a progress meeting in a simulated manner.

All four students took the "Operations Research" class in the university classes, and are familiar with schedule management, master schedule, critical path, and EVM. During the spring break, I created a master schedule for the first half and created the WBS from April to the end of July. Prepared by WBS style teachers and automatically generated parameters used in EVM such as SPI and CPI by updating WBS

The results of introducing the proposed method during the four months from April to July 2019 are shown below. In the university lab, Wi-fi environment, conference room, projector, laptop for all attendees, smartphone, and remote mouse, which is a voice input application, can be used. The environment required for the proposed method is in place. In that environment, according to the proposed method, we prepared the conference, made a report at the conference, and registered it in the task management ledger. The report from the manager to the upper manager was omitted.

### 4.1 Application result

Following the results of applying the proposed method by conducting a graduation examination seminar every week are reported.

#### 4.1.1 Transition of time required for seminar

At our laboratory seminar, first the faculty members will report the results of the past week and the schedule for this week. This seminar has been held regularly since January, but the only newly added measure since April is the method proposed in this paper. The following figure shows the change in the time required for the seminars that started in the second week of April and before the summer vacation.

#### 4.1.2 Transition of rework time

Documents used in the graduation seminar in the date folder on OneDrive must be read in advance. After the start of the

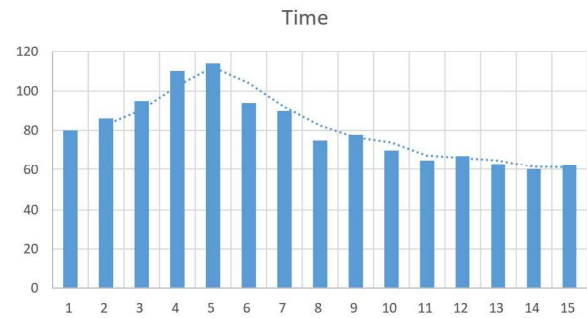


Figure 2: Transition of time for seminar

seminar, if the participants searched for the documents in the previous date folder, opened the file, and searched for the character string of MS-WORD, the time was recognized as rework and subsequently measured.

Next, the issues and homework that were newly found in the discussion at the seminar were entered in the issue management ledger. The following figure shows the changes in the rework time and number of tasks in the seminars that started from the second week of April and ended during the summer vacation.

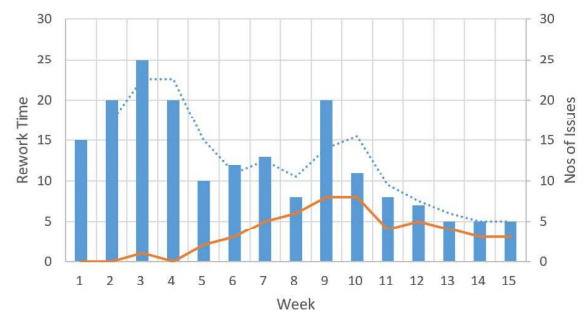


Figure 3: Transition of rework Time

#### 4.1.3 Voice Input Efficiency

The seminar in our laboratory was held in a paperless manner by using a projector at all times. During that time, the materials used during the seminar and the task management ledger were created by voice input by using a remote mouse.

Participants of the plenary session recorded the same content by handwriting, keyboard, and voice input, and examined the difference.

### 4.2 Evaluation of Issue solving

Here, we evaluate the elimination the cause of issue pointed at 2.2

#### 1. Occurrence of Rework

This system was launched in April, when the graduation seminar was held in this ceremony. At the beginning, the students were unfamiliar with the system,

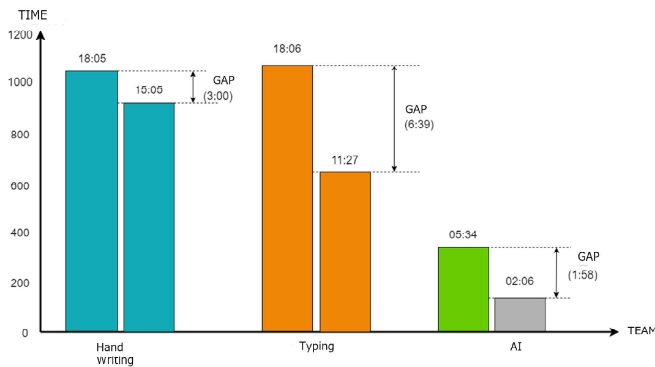


Figure 4: Voice Input Efficiency

and during the initial few weeks, the seminar hours increased. This was unavoidable regardless of the content. A lot of time is generally not required because only the results, problems of the past week, and work schedule for the present week need to be reported using one PowerPoint presentation. The problem was that the materials related to the seminar were stored in the past date folder, and the number of references gradually increased. However, as evident from the figure, the conference time and the reworking reduced from a particular point. Since the graduation examination seminar did not introduce any other measures besides the proposed method, it can be said that the effect of this proposal was confirmed.

## 2. Differences in Perception of Meetings

We decided to read the meeting materials prior to the meeting. In this study, the time to read the conference materials in advance is also counted as the conference holding time. On the contrary, in the initial few weeks, time increased. After that, the meeting time was also reduced and the awareness of the preparations for the meeting was revised.

## 3. Complicated Meeting Procedure

The actual time-consuming task related to the meeting was the creation of materials and minutes during the meeting. However, after introducing the proposed method, the materials during the meeting were created by voice input. Minutes transcript transcribes recording results. As the voice input and transcription are not perfect, subsequent rework will be required. However, the majority of laborious work has been improved.

## 5 Discussion

As shown in the three schemes of the proposed method, the improvement was confirmed in all cases. Although using voice input can be time-saving, it is difficult to predict the correction time for subsequent typographical errors. That is, the proposed method takes time to get used to.

However, for Office 365, which is used for everyday work, dictation by voice input has become possible from April 2020. As of June 2020, it is not an official release because it is a review version, but there is no existing functional problem. Apple, Google, Microsoft, and other IT giants offer voice input as well. Keyboard input will be reduced in the future, and voice input will subsequently increase. Accordingly, it can be concluded that being trained to use voice input from school days itself and learning to create minutes by transcription is an advantage.

## 6 Conclusion

In this study, we proposed the improvement of the conference process in companies and universities using IT and AI techniques.

The proposed method can also be applied to video conferences, which are expected to become considerably popular in the future. In the case of a video conference, everyone participated from the beginning via a PC and spoke via a voice cable. Recording can be done at any time using the function of the video conference and voices can be easily transcribed. It can also be applied for telecommuting and tele-work.

Currently, we are only considering university seminars. However, if an opportunity arises, we would like to evaluate the effectiveness of this proposed method by applying it to progress meetings at companies as well as other meetings.

## REFERENCES

- [1] ITmedia NEWS SPECIAL <https://www.itmedia.co.jp/news/articles/1709/29/news015.html>
- [2] Tomoo Inoue, et al, "Spatial Design for Integration of Face-to-Face and Video Meetings: HERMES Videoconferencing System", IEICE D2, PP.2482-2492, 1997
- [3] Hironori Tomobe, et al, "Discussion Media: Structuring and Browsing System for Discussion Contents", The 21st Annual Conference of the Japanese Society for Artificial Intelligence, PP. 1-4, 2007
- [4] Setsuko Hiramitsu, et al, "A Design of Face-to-Face Communication Support System Based on On-line Chat, SIG Human-Computer Interaction, IPSJ, Vol 2003, No.94, PP.7-12, 2003
- [5] Shintaro Sengoku, et al, Productivity management for intellectual interactions: an approach of transaction-based management, The Japan Society for Management Information, PP.30-33, 2009, Kazuo Yamamoto et al, Proposal for practical use of project meeting minutes, Project Management Society Vol. 2007, PP.363-367, 2007
- [6] Shin Mori, The Best Textbook of Outlook, Diamond Publishing, 2019

# Ad Hoc Transmission Algorithm for Low-Price Rice Field Server

Mikiko Sode Tanaka\*, Yuki Okumura\*\*, Sota Tatsumi\*\*, Shogo Ishii\*\*, Tatsuya Kochi\*\*

\* Global Information and Management, International College of Technology, Japan

\*\*Engineering Department, Kanazawa Institute of Technology, Japan

sode@neptune.kanazawa-it.ac.jp

**Abstract** - We develop a rice cultivation management system using field servers (FSs) to reduce the workload of farmers. The system realizes the possibility to manage environmental data for rice fields using sensors. The features of our FSs are reasonable price and mobility. To achieve the reasonable price, the accuracy of time synchronization is sacrificed. The system we have developed may not be able to obtain data due to rare accidents such as a car stopping next to an FS. This paper describes a new algorithm to solve the above problem. The proposed algorithm includes a data retransmission algorithm named simultaneous-transmission-type flooding algorithm. We also report the experimental results. This algorithm is robust for the rice cultivation management systems because it uses a robust resend algorithm in the FS. Therefore, it meets farmers' expectation of utilizing reasonable FSs.

**Keywords:** LoRa, IoT, Fail-safe, Multi-hop, Ad Hoc transmission algorithm, Rice, Time synchronization

## 1 INTRODUCTION

The agricultural working population has decreased by approximately 925,000 from 2010 to 2019 [1]. The situation in the field of agriculture has been dire. Hence, development of Internet of things (IoT) to support rice farming is desired.

We develop a rice cultivation management system using the field servers (FSs) to reduce the workload of farmers [2, 3, 4]. It is a system that can acquire environmental data of rice fields using a sensor, send the data to the master unit system using Low Power Wide Area (LPWA), and check it on the website. The system has a reasonable price; therefore, rice farmers can make a profit even if they introduce one FS for each rice field [2]. We set one slot as tens of seconds to allow for time error. Hence, the star method used in conventional LPWA communication was not practical because it required nearly 1 h to send data for 100 FSs to the master unit system. Therefore, we proposed a data collection algorithm to collect data within a short period of time [3]. Using this algorithm, it is possible to collect all the data in a few minutes, thus achieving a user's request within 5 min.

However, this system needs to be more reliable. Since the height of the FS is approximately 1 m, the developed system may not be able to obtain data owing to an accident such as signal interference due to a car stopping next to a FS. This is a very rare phenomenon, but the system must never make a transmission error. As shown in Figure 1, it is necessary to improve the reliability of retransmission by changing the transmission route.

Recently, the simultaneous-transmission-type flooding algorithm has been proposed as a data collection technology [5]. This method can build a stable and efficient sensor

network by repeated flooding without the use of routing. However, this method requires high-precision time synchronization and the apparatus becomes expensive. However, a low cost FS is desired in Japan. In addition, as with the star method, data collection requires a significant amount of time. Therefore, this method cannot be used to collect data from hundreds of FSs.

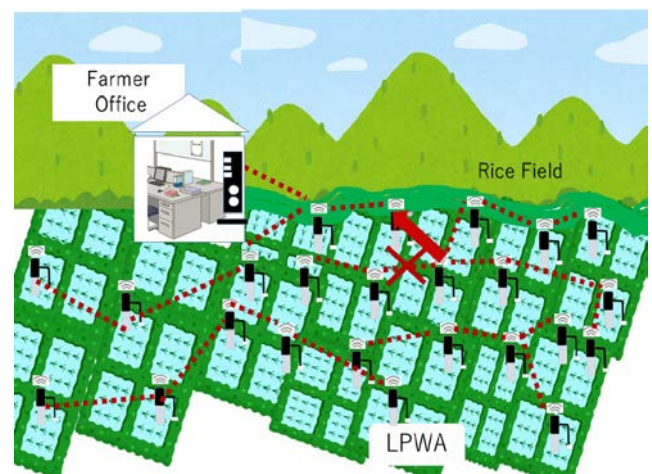


Figure 1 Transmission failure and transmission route change.

To achieve the reasonably priced FSs, the accuracy of time synchronization was sacrificed. For effective application to the rice field, it is necessary to have an algorithm that can tolerate time error, has high reliability of communication, and can collect data in a short period. In this paper, we propose an algorithm that collects data in a short period in an environment that accepts the time error.

The data are collected in a short period using the proposed data collection algorithm. Then, master unit system requests the FS that could not receive the data correctly to resend data using the simultaneous-transmission-type flooding algorithm. The FS that was requested to resend the data retransmits data using the simultaneous-transmission-type flooding algorithm. The simultaneous-transmission-type flooding algorithm is a very time-consuming method, but since the data collection algorithm shows almost no transmission error, this method was adopted because it is a useful method for transmitting data of a few FSs.

## 2 OVERVIEW OF RICE CULTIVATION MANAGEMENT SYSTEM

The rice cultivation management system monitors the water level in rice fields. The system is composed of the FS system, master unit system, and cloud service. Figure 2



shows the overall structure of the rice cultivation management system. The FS system is installed in the rice fields and accumulates sensor data for the water level. Further, the data are sent to the master unit system through the LoRa wireless network. The master unit system integrates the sensor data from the FS system and sends the data to the cloud service through the cellular telephone line or Wi-Fi. The water level can be checked on a mobile terminal via the cloud service. These services provide data to farmers serving as an alert regarding water levels, proposing a suitable work plan, preserving work records etc.

Communication between the FS system and the master unit system using LoRa is possible due to long-distance communication. LoRa has been estimated to have a practical communication distance of 3,000–4,000 m as shown via a basic communication characteristics survey conducted previously [1]. The rice field of Ishikawa prefecture was considered; the linear distance between the master unit system and the FS system was within 3,000 m. Hence, we adopted LoRa, which enables direct communication between the FS system and master unit system.

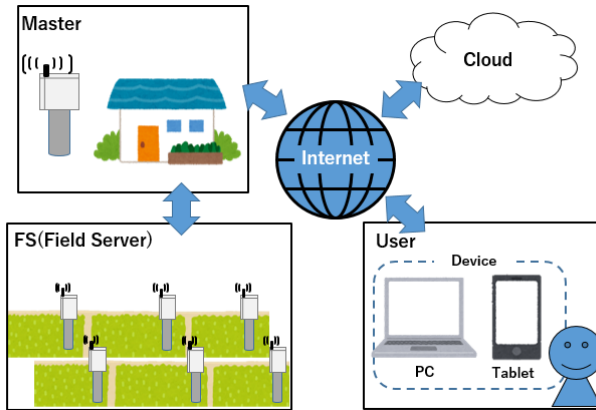


Figure 2 Rice cultivation management system.

### 3 DATA COLLECTION ALGORITHM

The data collection algorithm is explained [3]. An FS closest to the master unit system sends data to the master unit system. Other FSs create pairs and send data. The FSs that have sent data turn off the power. This is repeated until all the FSs are turned off. The data collection algorithm is detailed below.

- ① Turn on the power of each FS to start the servers.
- ② Send a transmission request by broadcasting from the master unit system to the FSs.
- ③ The FSs measure the sensor data.
- ④ The FS closest to the master unit system transmits the data to the master unit system. The remaining FSs establish a connection with each other using the shortest distance and transmit the data from the remote FS to the pair FSs.
- ⑤ Turn off the power of the FSs that have completed the transmission of data.

⑥ Repeat④ and ⑤ until all FSs transmit data and their power has been switched off.

Figure 3 shows an example of the six FSs. In step 1, FS 1 transmits the data to the master unit system, FS 2 transmits the data to FS 3, and FS 4 transmits the data to FS 5. In Step 2, FS 3 transmits the data received from FS 2 and the data held by it to the master unit system. FS 5 transmits the data received from the field server 4 and the data held by it to FS 6. In step 3, FS 6 transmits the data to the master unit system. In the case of six FSs, the direct method requires six steps; however, the data collection algorithm consists of three steps and is twice as fast.

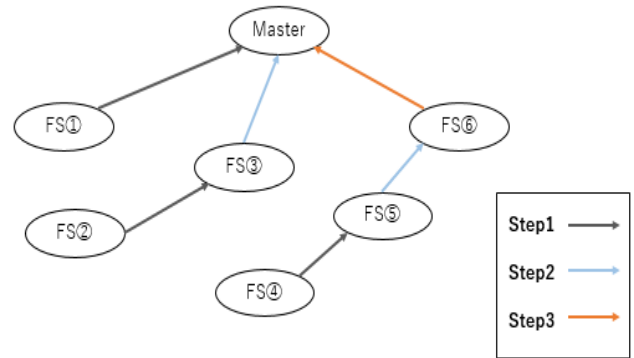


Figure 3 Example of transmission to the master unit system.

### 4 SIMULTANEOUS-TRANSMISSION-TYPE FLOODING ALGORITHM

The simultaneous-transmission-type flooding algorithm is explained in [4]. The first FS or master unit system broadcasts data to its range and powers down. Each FS that receives the data immediately broadcasts it and turns off the power. This is repeated until all the FSs are turned off.

Figure 4 shows an example of six FSs. In step 1, the master unit system broadcasts data. FSs 1, 2, and 3 receive the data. In Step 2, FS 1 broadcasts the data. Further, FSs 2 and 3 broadcast the data. FSs 4, 5, and 6 receive the data.

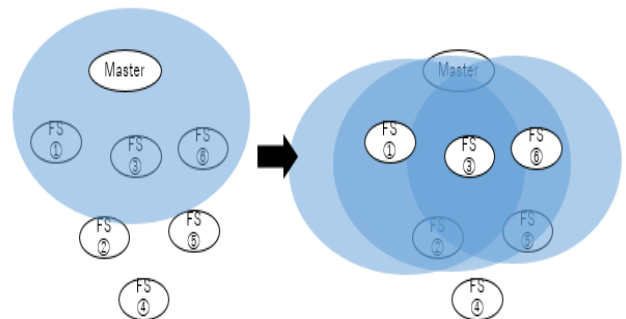


Figure 4 Hierarchy broadcasting from the master unit system.

Simultaneous-transmission-type flooding method increases the number of packet transmissions as compared with the routing method; however, as the sequence is repeated until all FSs receive and transmit the data, it is definitely the best way to broadcast data to all FSs. This method was adopted because all nodes need to obtain information on routing transmission, and it is necessary to reliably send the data of unsent nodes to the master unit system.

## 5 AD HOC TRANSMISSION ALGORITHM

Roads exist in the rice fields. Since the FS for rice field is approximately 1 m high, the transmission may not be performed correctly if a car is parked in the path of propagation. When the position of the antenna of the FS is half the height of the car, it may interfere with radio waves. As shown in Figure 5, when the FS and the office are close to each other, it is difficult for the radio waves to go around an obstacle; the radio waves may not reach the master unit system under the following conditions. Condition 1 refers to a car parked within 1 m of the FS antenna. Condition 2 refers to a car parked along the straight line connecting the FS and the office. To ensure transmission accuracy, we propose an ad hoc transmission algorithm to solve this problem.

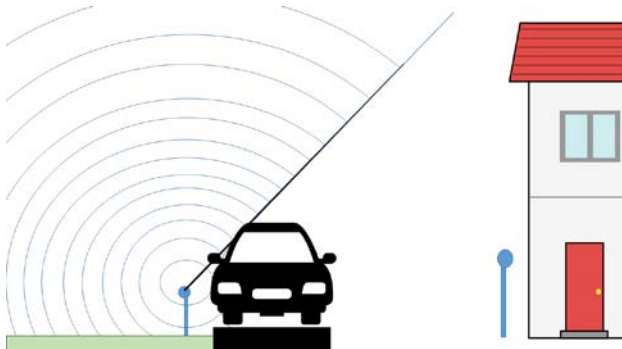


Figure 5 State of radio wave propagation.

The data collection algorithm collects data from all FSs in a short period. However, since the data collection algorithm transmits data through a fixed route, sometimes the data may not be transmitted correctly. In such cases, the data will have to be resent via the simultaneous-transmission-type flooding algorithm. After data collection, the master unit system transmits the retransmission scheduling information indicating whether all the data has been correctly received using the simultaneous-transmission-type flooding algorithm. When the retransmission scheduling information is empty, it indicates that all data have been transmitted correctly. When retransmission scheduling information is present, data are transmitted from the FSs to the master unit system using the simultaneous-transmission-type flooding algorithm according to the scheduling. The detailed ad hoc transmission algorithm is shown below.

//Ad hoc transmission algorithm

```

1: Collect data from all FSs using the data
   collection algorithm
2: While(){
3:   if (Was the master unit system able to collect data from
       all FSs?)
4:     then{
5:       Master unit system transmit the retransmission
       scheduling information by using the simultaneous-
       transmission-type flooding algorithm.
6:       Break;}
7:     else{
8:       Master unit system transmit the retransmission
       scheduling information by using the simultaneous-
       transmission-type flooding algorithm.
9:       Data are transmitted from the FSs to the master
       unit system using the simultaneous-transmission-type
       flooding algorithm according to the scheduling.
       }

```

## 6 EXPERIMENTAL RESULT

### 6.1 Experimental Environment

We aim to develop a low-cost FS. Hence, the time error of the microcomputer is accepted and an expensive element such as a special crystal is not used. In the proposed method, the time is corrected once every hour when data are transmitted. Figure 6 shows the results of measuring the time error of the selected microcomputer ATMEGA328P-PU. The average time error is 3 s, the variance is 9, and the maximum error width is 7 s.

The proposed ad hoc transmission algorithm was tested using a low-cost FS with time error. Since this is an operation confirmation experiment, the experiment was conducted indoors. We performed the experiment without an antenna. One master unit system and six FSs were used. IoT devices such as FSs are used outdoors. As there is no power supply, conventional measuring instruments cannot be used. We need to use a battery-powered measuring instrument. However, as such a measuring instrument is not available, we developed a battery-powered measuring instrument using a current sensor module called INA219 for current and time measurements [6].

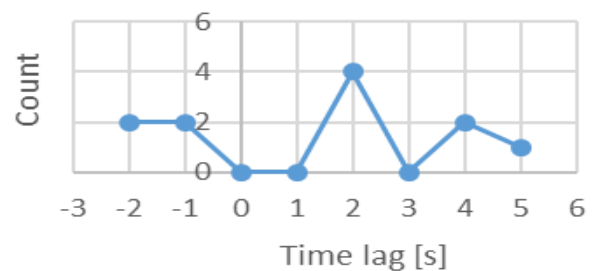


Figure 6 Microcomputer time error.

## 6.2 Experimental Results of the Data Collection Algorithm

The experimental results of the data collection algorithm are shown in Figure 7. Figure 7(a) shows an example of the six FSs, while Figure 7(b) shows the sequence diagram of the example. In addition, Figure 7(c) shows the current waveform of FSs 4, 5 and 6. Look at the circled area of Fig. 7(C). The field server 6 is up for a few seconds, then the field server 5 is up, and a few seconds later, the field server 4 is up. In this way, our device has a time error, so it cannot start at the same time. Similarly, the timing to turn off the power is when the data is sent, so you can confirm that the power has not been turned off at the same time.

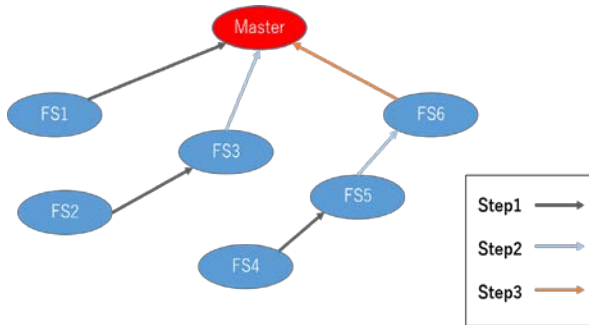


Figure 7(a) Scheduling results with six FSs.

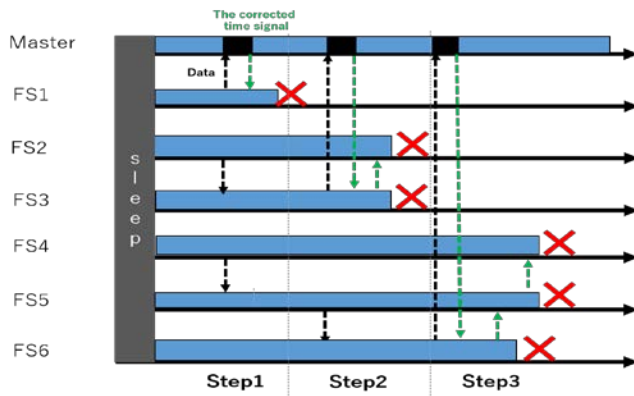


Figure 7(b) Sequence diagram of Figure 7(a).

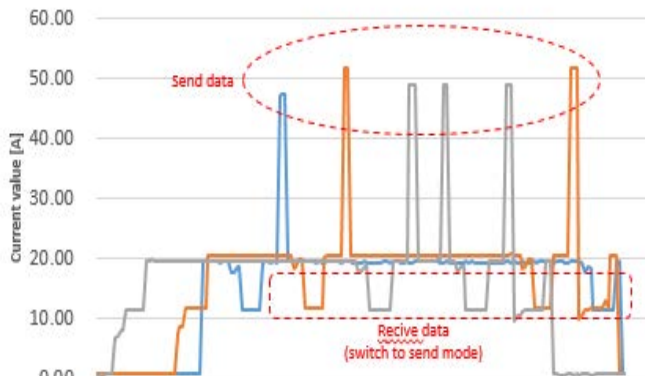


Figure 7(c) Current measurement result of Figure 7(a).

The current waveform indicates that the transmitted current is approximately 50 mA. Data are transmitted in the

following order: FS 4, FS 5, and FS 6. Using this process, the data of FSs 4, 5, and 6 are sent to the master unit system. Next, the time correction signal from the base unit are sent to FS 6, FS 6 sends the time correction signal to FS 5, and then FS 5 sends the time correction signal to FS 4. Thus, we can confirm that processing is carried out correctly.

## 6.3 Experimental Results of Simultaneous-Transmission-Type Flooding Algorithm

The experimental results of the simultaneous-transmission-type flooding algorithm are shown in Figure 8. Figure 8 shows measurement results for one round. In step 1, the master unit system transmits the scheduling information by broadcasting. FSs 1, 2, and 3 receive the data and change the mode from reception mode to transmission mode, which requires 10 s. In step 2, since there are individual differences between the FSs, in this example, FS 1 transmits data first. Since the transmission cannot be performed at the same time, the other FSs wait for the transmission to end. The data sent by the FS 1 are received by FS 6. FS 6 changes the mode to the reception mode. FS 2 sends data, which are received by FS 5. Similarly, FS 3 sends data, which are received by FS 4. FSs 4, 5, and 6 also transmit data. This process seems unnecessary, but it is necessary because there is no way to check if all FSs have received the data. Whenever an FS receives data, it sends the data to complete the processing.

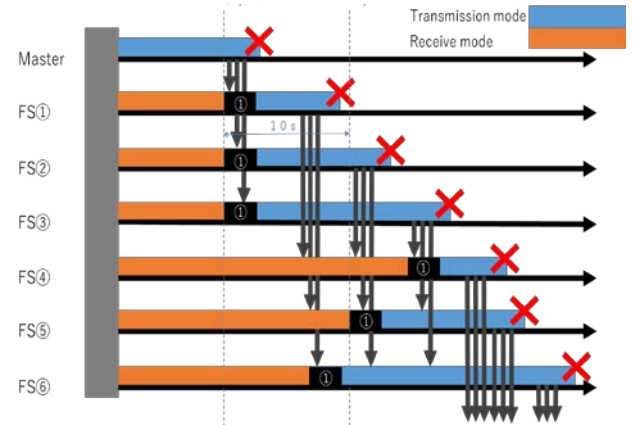


Figure 8 Sequence diagram of the simultaneous-transmission-type flooding algorithm.

## 6.4 Experimental Results of the Ad Hoc Transmission Algorithm

We explain the channel setting for the ad hoc transmission algorithm. The transmission channel is set as shown in Table 1 such that no collision occurs. The simultaneous-transmission-type flooding algorithm used FS channel setting mode 1. The data collection algorithm used FS channel setting mode 2.

An experiment was conducted to confirm the operation of the proposed algorithm that sends the data of all FSs to the master unit system. Since data transmission errors rarely occur, we simulated an environment where data transmission errors occur intentionally and conducted experiments. Two examples are shown to display the working of the algorithm.



Table 1 FS Channel Settings

	#Node	Mode1 #Channel	Mode2 #channel
Master Unit	1	2	2
FS①	2	2	2
FS②	3	2	2
FS③	4	2	2
FS④	5	2	2
FS⑤	6	2	3
FS⑥	7	2	4

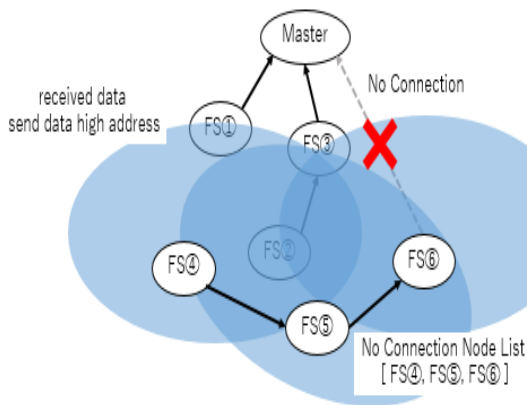


Figure 9(a) Test example 1.

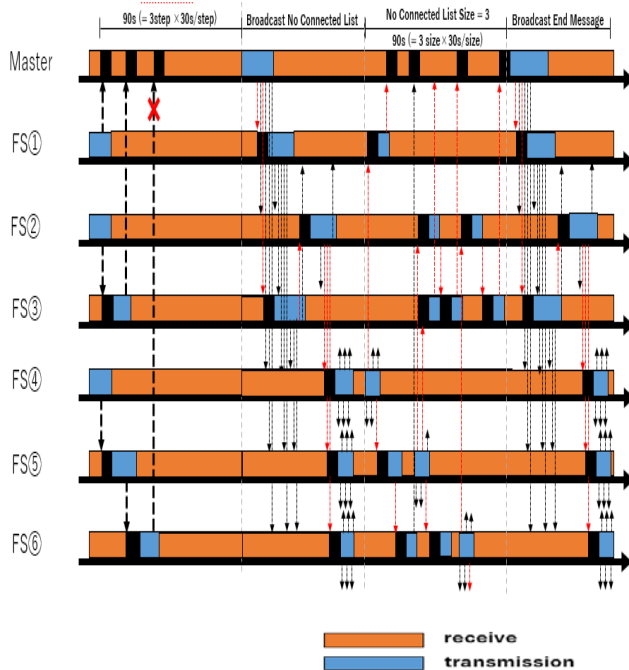


Figure 9(b) Sequence diagram of test example 1.

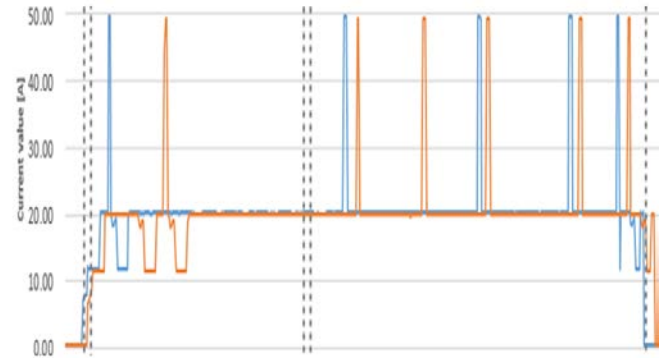


Figure 9(c) Current measurement results of test example 1.

An experimental result of an example where transmission from FS 6 to the master unit system is disabled is shown in Figure 9. Figure 9(a) shows the operation image of the data collection algorithm. In step 1, FS 1 sends data to the master unit system, FS 2 sends data to FS 3, and FS 4 sends data to FS 5. In step 2, FS 3 sends data to the master unit system and FS 5 sends data to FS 6. In step 3, FS 6 sends data to the master unit system. However, communication is not possible due to the large distance between FS 6 and the master unit system.

The master unit system determines that the communication is not complete, and broadcasts the retransmission scheduling information of FS 6 for which communication is not completed, using the simultaneous-transmission-type flooding algorithm. In the next step, FS 6 retransmits data using the simultaneous-transmission-type flooding algorithm. The master unit system rechecks whether all data have been received. When the master unit system determines that the communication has been completed, it broadcasts the retransmission scheduling information indicating that the communication has been completed.

Figure 9(c) shows the measurement results of the current waveform during operation in Figure 9(b). We focused on FS 1 and FS 3, and measured the current. The current at the time of transmission is approximately 50 mA. We compare the position of the transmission on Figure 9(c) to the blue box position showing transmission in Figure 9(b). Thus, we can confirm that the transmission position is the same.

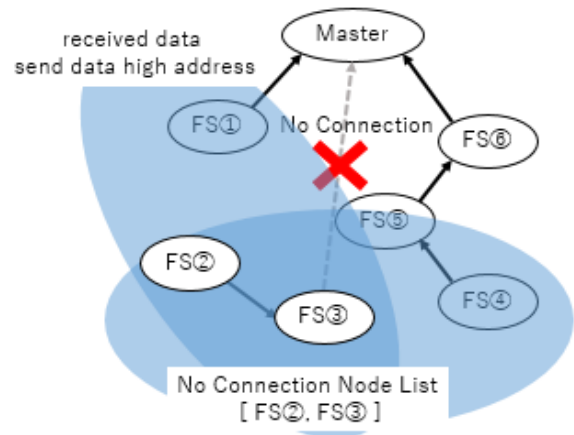


Figure 10(a) Test example 2.

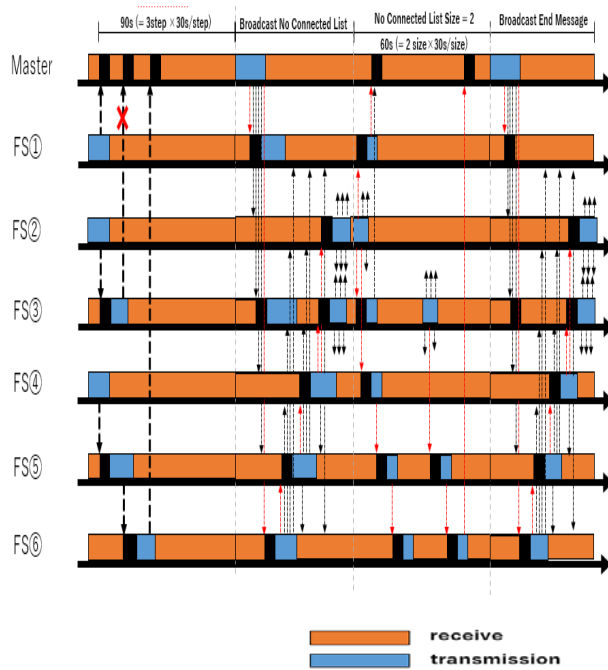


Figure 10(b) Sequence diagram of test example 2.

The process in which all FSs acquire sensor data and collect the data in the master unit system is called one round. One round of processing is performed every hour. The second example is the measurement results for one round, which are shown in Figure 10(a). First, communication is performed according to the scheduling information. In step 1, FS 1 sends data to the master unit system, FS 2 sends data to FS 3, and FS 4 sends data to FS 5. In step 2, FS 5 sends data to FS 6 and FS 3 sends data to the master unit system. However, communication is not possible due to the large distance between FS 3 and the master unit system. In step 3, FS 6 sends data to the master unit system. The master unit system determines that the communication is not complete, and broadcasts the retransmission scheduling information of the FS for which communication is not completed in hierarchy, and shares the information with all FSs. Each FS compares the node information of the FSs that have not completed the communication with their own node number, and if the codes match, broadcasts data in hierarchy. When the transmission described in the retransmission scheduling information is complete, the master unit system rechecks whether has all data have been received. When the master unit system determines that the communication has been completed, it broadcasts retransmission scheduling information indicating that the communication has been completed.

In this manner, even if the FS cannot receive the data due to some error, it can receive data from other nodes considering the redundant transmission. Therefore, it is a robust algorithm.

The data collection algorithm requires approximately  $\log 2n$  steps when the number of FSs is  $n$  [7]. The master unit system sends the retransmission scheduling information to FSs in one step; each unsent FS sends the data to the master unit system in one step. Therefore, the proposed algorithm is a useful technique in situations where there are few retransmissions.

## 7 CONCLUSION

Low cost FSs have poor time synchronization accuracy. Therefore, an algorithm that allows for a time error was required. We proposed an algorithm that works with low cost FSs and could collect data in a short period. However, this method was vulnerable to data transfer fail due to an accident. In this paper, we proposed an algorithm to improve the reliability of data transfer. The conventional method utilized a simultaneous-transmission-type flooding technology. This technology requires highly accurate time synchronization; however, by avoiding this perfect simultaneous transmission, it is possible to successfully operate with a low cost FS.

As a result of the experiment, we have confirmed that all data of FSs can be sent. This protocol is robust for the rice cultivation management systems because the FS is stable. Therefore, it meets farmers' expectation to utilize a reasonable FS.

## ACKNOWLEDGMENT

This research is supported by 2018 Ishikawa commercialization promotion support project.

## REFERENCES

- [1] Ministry of Agriculture, Forestry and Fisheries, <https://www.maff.go.jp/j/tokei/sihyo/data/08.html>
- [2] Koichi Tanaka, Mikiko Sode, Masakatsu Nishigaki, Tadanoro Mizuno, "A Study on Time Synchronization Method for Field Servers for Rice Cultivation," International Journal of Informatics Society (IJIS) (2019).
- [3] Tatsuya Kochi, Sota Tatsumi, Yuki Okumura, Shogo Ishii, Mikiko Sode Tanaka "Experimental Results of Aggregated data communication algorithm for rice cultivation," IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech 2020), 2020/3/10-12.
- [4] Shogo Ishii, Tatsuya Kochi, Sota Tatsumi, Yuki Okumura, Mikiko Sode Tanaka, "New Field Server Addition Method for Low-Price Rice Cultivation Management System," IEEE International Conference on Computer Communication and the Internet (ICCCI 2020), 26th–29th June 2020.
- [5] F. Ferrari, M. Zimmerling, L. Thiele, and O. Saukh, "Efficient network flooding and time synchronization with glossy," Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), pp. 73–84, 2011.
- [6] Yumeto Kojima, Mikiko Sode Tanaka, "Current measuring instrument for field management field server using LoRa," 2018 IEICE Society Conference, 2018/9/11-14.
- [7] Mikiko Sode Tanaka, "Proposal of IoT Communication Method for the Rice Field," International Journal of Informatics Society (IJIS) (2020)

# Java Model Checking: Improvement of the Understanding of Counterexample

Chellet Marwan Bernard Hassan, Shinpei Ogata, and Kozo Okano

Shinshu University, Nagano, Japan  
19w2074f@shinshu-u.ac.jp,{okano,ogata}@cs.shinshu-u.ac.jp

**Abstract** - The reliability factor during the software creation process is a very important point. When creating software it's not rare to find design errors that can lead to bugs. A verification method called model checking is used to resolve these types of errors. The principle is to verify if the software or a given fragment of program code satisfies important properties which can give bugs if they are violated. If one of the properties is violated, it will provide a counterexample which helps to localize the issue. Counterexamples are often given in a sequence of machine instructions that can make it difficult to understand. This paper will offer a new way to treat counterexamples and makes it more useful and convenient. Model checkers for C language already have some solutions but nothing is made for Java language which is still one of the most used for software programming. This research focuses on a way to translate a counterexample from a model checker for Java language into Java code which is easier to understand. Currently, the counterexample is given in Java Bytecode instructions. It will be translated into lines of Java code with a library such as Procyon. The counterexample also gives a line number from the source code of every instruction. Each of these lines will be replaced by the corresponding translated code. The interest of this code could be to make it executable in order to directly locate the error by using a debugger for instance.

**Keywords:** Model checking, Counterexample, Java, Simulation, Verification

## 1 Introduction

In the last few years, the software reliability became a real challenge for the developers. Effectively, when we create a software it's not rare to find design errors which can lead us to big bugs and can have real consequence. A lot of testing methods are already used in order to resolve this type of bugs; however, it's not easy to get full coverage and it's using the system itself to find errors.

Another way to increase the correctness and reliability of software is to use a verification method called "model checking". It's an automated method which simulates the system behavior by using a finite-state model instead of the real system. It permits us to find complex bugs that the test methods cannot find and have better coverage. However compared to test methods, it also has disadvantages.

The last big disadvantage of model checking is the given counterexample for a violated property. Instead of the test method in which the output is usually given in an understandable way by using code, counterexamples are usually sequences of transitions which leads to the error state from the

initial state. For the software model checking which inputs program code instead of state machines, the counterexamples are usually represented in a sequence of low-level machine codes. This is caused by some limitations on the current software model checking tools. It makes it difficult to understand for people who are not experts in machine language and can be a problem in order to localize the bug. This paper introduces a way to improve the comprehensibility and accessibility of model checking by translating the machine language counterexample into Java code and using one of the most effective java model checker called JBMCC[1].

This paper is organized as follows: Section 2 introduces the concept and principle of model checking, the model checker we used, and the part of model checking we will focus on. Section 3 explains our proposed method for this paper. Section 4 provides the experiment we did and the result we got. Section 5 talks about the remarks we got during the creation of the proposed method. Section 6 gives us the future work which should be done to follow up on this solution. Section 7 provides all the existing works related to this solution and finally, Section 8 summarizes this paper.

## 2 Model checking

In this section, we define the principle of model checking and how it works. After that, we explain our java model checker choice. The last part focuses on the counterexample which is the central part of this paper.

### 2.1 Principle of model checking

Model checking has already proven itself as a verification method for software. The goal of this method is to find violated properties into a system translated into an abstract model. The model is represented as a transition system. It is an oriented graph: a vertex represents a state of the system and each arc represents a transition, that is to say, a possible evolution of the system from a given state to another state. This method works in 3 phases: The first one is the modeling phase, it takes the real system and translate it into an abstract model. The system became a set of states and transitions. States give information about the program such as variable values. Transitions are here to describes how the system works. It also formalizes the properties before being verified. The second phase is the running one, the model checker will check the possible paths of the abstract model to find a violated property. The last one is the analysis phase, this part is the one we are interested in. There is three different possible output from the running phase. The first output is when a property is violated, the model checker gives a counterexample. The model checking principle[15] and its three phases

are described in Figure 1. It aims to help developers to localize the bug by giving the system behavior which giving us the error. The second one is when all the properties are verified and not violated, the model checker tells us the program has been verified and it didn't find bugs. The third output is when a software is too complex and the abstract model becomes too big, the model checker can't handle it and run out of memory.

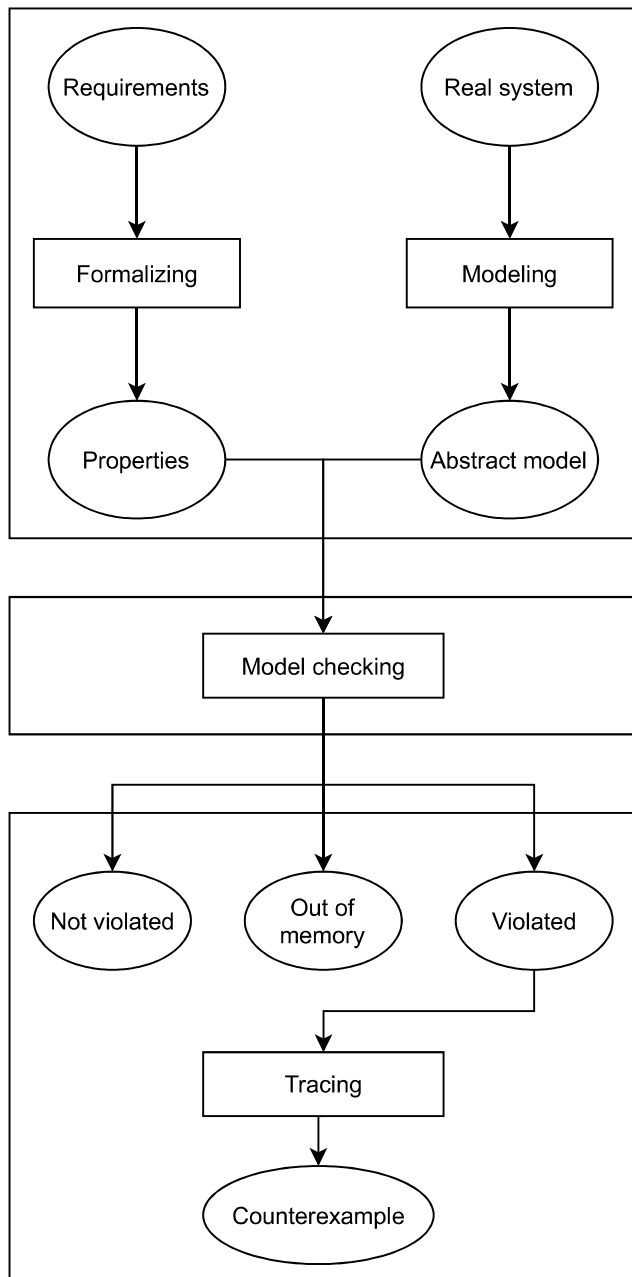


Figure 1: Principle of model checking

The model checker we gonna use is a bounded one. Bounded model checking method[2] is a way to use model checking by traversing a Finite-State Machines with a fixed number of steps  $k$  a check if a property is violated in this bound. More the value of  $k$  is big and more you have a chance to find a violation. The principal goal is to be faster and per-

formant by giving a limit of steps.

## 2.2 Model checking with JBMC

JBMC is a bounded java model checker developed in C++ and based on the C model checker called CBMC[4]. This model checker takes a Java Bytecode program in input. The core of the model checker is managed by the framework CPROVER[16]. It's a bounded model checker so the last input will be the number of bounds,  $k$  we want to do.

JBMC works in 4 different steps: The first step is to parse the java bytecode into a parse tree which is correspond to the translation into a Finite-State Model step. The second one is to translate into the CPROVER control-flow graph representation which is called GOTO program. The goal of this step is to simplify the Java Bytecode representation and make it easier to analyze. The third one corresponds to the running phase of the model checker which gonna analyze the model and check properties. The last step is the SAT solver one which tells us if a property is violated for a given bound,  $k$ , and gives us a counterexample if it's the case.

Figure 2 illustrates a simple example of a Null Pointer Exception error which can be found by JBMC. It's a really simple program that just prints values of an array of strings. The variable "v" takes a random value between 1 and 2. If the value is 1, then only the first value of the array "tab" will be printed but if the value is 2, then it will also print the second value which is null and throw a Null Pointer Exception.

```

import java.util.Random;

public class NullTest
{
    public static void main(String [] args)
    {
        String [] tab = {"a", null};
        Random r = new Random();
        int v = r.nextInt((2-1)+1)+1;
        for (int i=0; i<v; i++)
        {
            System.out.println(tab[i].length());
        }
    }
}

```

Figure 2: Code with a Null Pointer Exception error

## 2.3 Counterexample

When a property is violated a counterexample is given. This output aims to help developers to localize the bug by giving the behavior which leads us to the bug such as variables values. As we said before, JBMC uses CPROVER which produces the counterexample. The trace produced in Figure 3 is the output given by the model checker after the verification of the code showed in Figure 2.

```

State 84 function java:java.lang.System.<clinit>:(V thread 0
-----
dynamic_object16={.@class_identifier="java:java.io.PrintStream" } { ? ? }
State 91 file NullTest.java function NullTest.main(java.lang.String[]) line 12 thread 0
-----
this=&a
(00000000 00010001 00000000 00000000 00000000 00000000 00000000 00000000)
State 99 file NullTest.java function NullTest.main(java.lang.String[]) line 12 thread 0
-----
this=&dynamic_object16
(00000000 00010011 00000000 00000000 00000000 00000000 00000000 00000000)
State 100 file NullTest.java function NullTest.main(java.lang.String[]) line 12 thread 0
-----
stub_ignored_arg1=1 (00000000 00000000 00000000 00000001)
State 102 file NullTest.java function NullTest.main(java.lang.String[]) line 10 thread 0
-----
anonlocal::4i=1 (00000000 00000000 00000000 00000001)

Violated property:
file NullTest.java function NullTest.main(java.lang.String[]) line 12 thread 0
Null pointer check
!((struct java.lang.String *)((struct java::array[reference] *)anonlocal::1a)-
>data[anonlocal::4i] == null)

```

Figure 3: Trace of counterexample from JBMC

We can easily see what is the violated property but it's difficult to understand how the model checker went into this violation and which path should we take to get the same behavior. Every states in the counterexample represent one instruction such as variable value but this instruction is given by the framework CPROVER which is initially made for C and C++ programs. The problem is the difficulty to understand these instructions and try to see what is the java equivalence.

### 3 Proposed method

This section describes the proposed method which is a way to translate the counterexample given by JBMC into Java code. Figure 4 shows the process of the whole system. The first step is just the transformation of java code into Java Bytecode before using the model checker. The second one is the running of the model checker, if a property is violated we are going to step three. The third step is the treatment of the counterexample and this paper focuses here. The last one is the report of the output as a java code after the translation of the counterexample.

#### 3.1 Step 1: Analyse of counterexample

This solution is based on a transformation algorithm. The first part will be to add all the new instructions from the counterexample such as variable definition into the source code. To do this we have to locate the line on which the basic instruction of the source file has changed. After that, we take all CPROVER instructions from the counterexample related to this line and translate it into only one line of java code. After getting all these new lines of codes, the focus is on the violated property. When the property is not from a user assertion, it creates a new line of code with an assert just before the line where the property is violated.

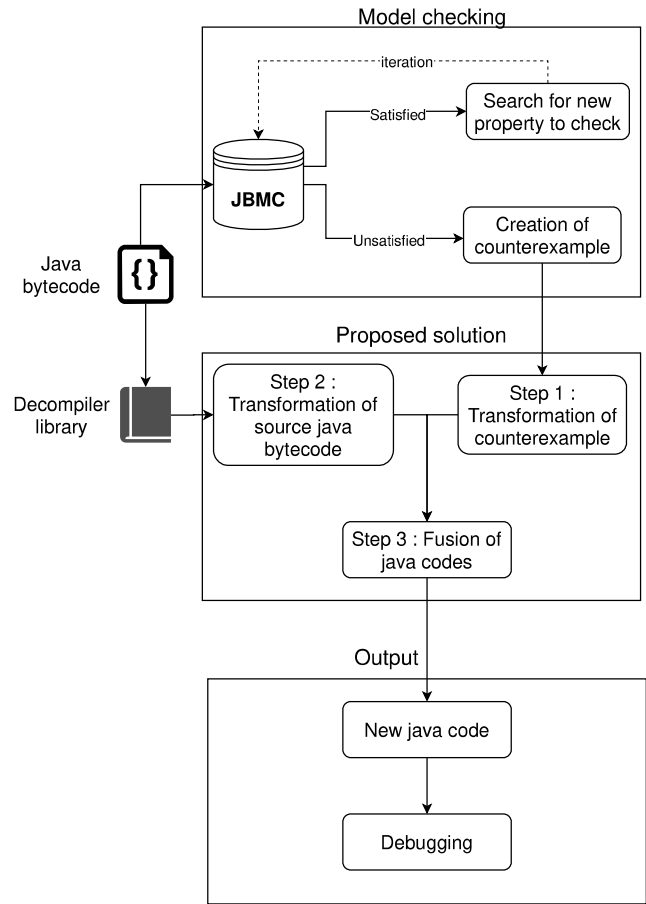


Figure 4: Proposed method process

#### 3.2 Step 2: Transformation of Bytecode source file into java

The second part of this solution is to transform the source file which is actually in Bytecode into java code again then it can be used after for debugging. To do that there is already some java library which transform bytecode into java. There are some good libraries such as Procyon but we choose CFR which is not optimizing the code by some technique such as deleting dead code or using variable propagation. It permits the user to find the same code he wrote before and to localize the error more easily.

#### 3.3 Step 3: Add new java lines into the transformed source file

This last step is the fusion of the two previous ones. The goal was to get a java code which reproduce the same behavior of the software when the model checker found the violation of property. The variables names from the Bytecode and the counterexample are not the same and one of the difficulty is to be sure to change the good variable by using the line given in the counterexample. Figure 5 represent the final result of the proposed method for the Figure 2.

Table 1: The Results of the Experiments

	Execution Time	New Assertion	New Instructions	Fusion
Null Pointer	1.03s	OK	OK	OK
Index Out of Bound	0.50s	OK	OK	OK
Division By 0	0.95s	OK	OK	OK
User Assertion	0.88s	NONE	OK	OK

```

import java.util.Random;

public class NullTest
{
    public static void main(String [] args
    )
    {
        String [] tab = {"a", null};
        Random r = new Random();
        int v = 2;
        for (int i=0; i<v; i++)
        {
            assert (tab[i] != null);
            System.out.println(tab[i].length()
            );
        }
    }
}

```

Figure 5: New code generated by the solution

```

import java.util.Random;

public class IndexTest
{
    public static void main(String [] args
    )
    {
        int [] test = {1};
        Random r = new Random();
        int v = r.nextInt((2-1)+1)+1;
        for (int i=0; i<v; i++)
        {
            System.out.println(test[i]);
        }
    }
}

```

Figure 6: Source code of an Index Out of Bounds Exception

```

import java.util.Random;

public class IndexTest
{
    public static void main(String [] args
    )
    {
        int [] test = {1};
        Random r = new Random();
        int v = 2;
        for (int i=0; i<v; i++)
        {
            assert (test.length >1)
            System.out.println(test[i]);
        }
    }
}

```

Figure 7: New generated code of an Index Out of Bounds Exception

## 4 Experiment and results

To see the efficiency of the proposed solution, we tried it with different simple programs such as Figure 4 with different types of errors which JBMC can find a counterexample. We get one program per error, Null Pointer Exception, Index Out Of Bounds Exception, Division By Zero Exception, and a user assertion. For this experiment we choose a number of bounds  $k = 5$ . Table 1 shows us the result of the experiment for each program step by step. Figure 6 and Figure 7 are an example of the experiment result with an Index Out of Bound Exception.

All the cases were successful and this experiment shows us it is possible to translate a counterexample into Java code. This experiment was made with very simple programs but it's a good start for the understanding of counterexample.

## 5 Discussion

Even with this solution, the understanding of counterexample in java model checker seems to be still difficult. It's open a way for researchers to think about model checking in java and how to improve it. During this research, we saw that the Java model checking don't have a lot of solutions to be improved such as C model checking but it still one of the most used language for software development. This step of making counterexample more understandable is an important one in

order to make model checker more convenient and useful for casual use.

## 6 Future work

This solution can still be improved to be more efficient. One of the best ways can be to use AI techniques which can directly predict the kind of property violated and how to resolve it. An interesting one can be deep learning by using neural networks but to use this technique we need a large amount of dataset before starting. The next step can be to make the newly generated code more useful by translating it into a JUnit test which can be more efficient to find bugs.

## 7 Related work

Try to improve counterexample generation in model checking is not something new ([6], [8], [9], [10], [11], [12], [13]). One of the first approach about executable counterexample was made by Dirk Beyer [5]. This paper is inspired by the work of Understanding Programming Bugs in ANSI-C Software Using Bounded Model Checking Counter-Examples [14] which is doing a counterexample simplification for AINSI-C software by using ESBMC model checker which also uses the CPROVER framework. All the understanding of the model checking was based on the book Principles of Model Checking [15].

## 8 Conclusion

To conclude, this solution was created to help java developers who are not experts in verification methods to use model checking easily. It was tested with simple and small programs and has to prove the efficiency of complex software. It can still be improved and open doors for future works such as using AI techniques to make it more automatic and efficient. Model checking and more specifically in java is still under-used because of the complexity of use and the inconvenience. This solution can be a beginning for model checking to become a common tool for software developers.

## Acknowledgement

The research is being partially conducted as Grant-in-Aid for Scientific Research A (19H01102).

## REFERENCES

- [1] Cordeiro, L., Kroening, D., and Schrammel, P: "JBMC: Bounded Model Checking for Java Bytecode," In: Beyer D., Huisman M., Kordon F., Steffen B. (eds) Tools and Algorithms for the Construction and Analysis of Systems. TACAS 2019, Lecture Notes in Computer Science, vol. 11429, pp. 219-223. (2019)
- [2] Biere, A., Cimatti, A., Clarke, E.M., Strichman, O., and Zhu, Y: "Bounded model checking," In : Advances in computers, vol. 58, no 11, pp. 117-148. (2003)
- [3] Strichman, Ofer: "Tuning SAT Checkers for Bounded Model Checking," In : Computer Aided Verification. CAV 2000, Lecture Notes in Computer Science, vol. 1865., pp. 480-494. (2000)
- [4] Clarke, E., Kroening, D., and Lerda, F: "A tool for checking ANSI-c programs," In: International Conference on Tools and Algorithms for the Construction and Analysis of Systems. TACAS 2004, Lecture Notes in Computer Science, vol. 2988, pp. 168-176. (2004)
- [5] Dirk Beyer, Adam J. Chlipala, Thomas A. Henzinger, Ranjit Jhala, and Rupak Majumdar: "Generating Tests from Counterexamples," In : Proceedings of the 26th International Conference on Software Engineering. ICSE 04, IEEE Computer Society, pp. 326-335. (2004)
- [6] Shankar, N., and Sorea, M: "Counterexample-driven model checking," In : Technical Report SRI-CSL-03-04, SRI International Computer Science Laboratory (2003)
- [7] Gennari, J., Gurfinkel, A., Kahsai, T., Navas, J. A., and Schwartz, E. J.: "Executable counterexamples in software model checking," In : Working Conference on Verified Software: Theories, Tools, and Experiments. pp. 17-37. (2018)
- [8] Seghir M. N., and Kroening D: "A visual studio plug-in for CProver," In : 2013 3rd International Workshop on Developing Tools as Plug-Ins. TOPI, pp. 43-48. (2013)
- [9] Müller P., and Ruskiewicz J. N.: "Using Debuggers to Understand Failed Verification Attempts," In : Formal Methods - 17th International Symposium on Formal Methods. FM 2011, pp. 73-87. (2011)
- [10] Ball, Thomas., Naik, Mayur., and Rajamani, Sriram: "From Symptom to Cause: Localizing Errors in Counterexample Traces," In : Conference Record of the Annual ACM Symposium on Principles of Programming Languages, pp. 97-105. (2003)
- [11] Groce, A., Kroening, D., and Lerda, F: "Understanding counterexamples with explain," In : International Conference on Computer Aided Verification. CAV 2004, Lecture Notes in Computer Science, vol. 3114, pp. 453-456. (2004)
- [12] Kroening, Daniel., Groce, Alex., and Clarke, Edmund: "Counterexample Guided Abstraction Refinement Via Program Execution," In : Formal Methods and Software Engineering. ICFEM 2004, Lecture Notes in Computer Science, vol. 3308, pp. 224-238. (2004)
- [13] Rustan, K., Leino M., Millstein, Todd ., and Saxe, James B: "Generating error traces from verification-condition counterexamples," In : Science of Computer Programming, Vol. 55, pp. 209-226. (2005)
- [14] Rocha, Herbert., Barreto, Raimundo., Cordeiro, Lucas., and Neto, Arilo: "Understanding Programming Bugs in ANSI-C Software Using Bounded Model Checking Counter-Examples," In: International Conference on Integrated Formal Methods. IFM 2012, pp. 128-142. (2012)
- [15] Baier, Christel. and Katoen, Joost-Pieter: "Principles of Model Checking," MIT Press. (2008)
- [16] Kroening, D., and Clarke, E.: "The CPROVER User Manual," <https://www.cprover.org/cbmc/doc/manual.pdf> (2001)





# Enhancing RTK-GNSS Infrastructure in Snowy and Mountain Region Through Rule-Based Base Station Assignment Approach

Bhagawan Rokaha\*, Bishnu Prasad Gautam\*\*, and Tomoya Kitani\*

\* Graduate School of Integrated Science and Technology, Shizuoka University, Japan

\*\* Department of Economic Informatics, Kanazawa Gakuen University, Japan  
b-rokaha@kitanilab.org, gautam@kanazawa-gu.ac.jp, t-kitani@kitanilab.org

**Abstract** – RTK-GNSS is a promising technique, where the base station plays an essential role, is widely used in various applications requiring precise positioning. However, the base station correction signals are often interrupted, delayed, or discontinuous in mountain and snowy regions due to snow accumulation, single base-line area, and multipath errors. It has been therefore felt necessary of a reliable RTK-GNSS infrastructure that can ensure the continuity and reliability of the base station data as well as enhance the overall performance of the positioning system. In this paper, we address these issues of RTK-GNSS by implementing two major components. The first component is a web-based real-time monitoring system that is developed in order to supervise the state of all surrounding base stations. The second component is a processing component, which is an algorithm specially designed to assign the most favorable base station from the list of multiple base stations. To achieve precise positioning, we proposed a rule-based base station assignment algorithm. In this approach, when the running base station has some problems or the user moves out of the base-line area, the next available yet the optimal base station is assigned dynamically. Experimental results demonstrate that the proposed approach can maintain the rover receiver positioning accuracy within the centimeter-level even after the base station handover. By combining these two components' results, we were able to improve the reliability of RTK-GNSS positioning in challenging environments through continuous monitoring and by providing users with the most favorable base station.

**Keywords:** RTK-GNSS, Base station assignment, Web-based monitoring, Reliable infrastructure

## 1 INTRODUCTION

Long-term stable, reliable, and highly precise positioning and navigation functionalities are mandatory in many applications, including autonomous vehicles, precision agriculture, weather forecasting, drone, Geographical Information Systems (GIS). In the conventional single point positioning system, the user's position is instantly determined using a pseudo-range between each satellite and the user's receiver, where four or more satellites are needed. However, the positioning accuracy ranges from 10m to 30m, as various factors added as an error in the GPS observation[1]. Therefore, many technologies are developed to enhance positioning accuracy. One of the famous differential positioning systems that emerged since a few decades ago that provides centimeter-level accuracy is a

Real-Time Kinematics-Global Navigation Satellite System (RTK-GNSS)[2]. In this system, a higher resolution distance information called a phase pseudorange is used instead of the code pseudorange. As shown in Figure 1, the precise position of a user or rover station is calculated with receiving measurement signals from satellites and the correction signal from the base or reference station. The precise position, called a fixed position, is within a few centimeters. However, the positioning accuracy of RTK-GNSS is limited by different errors, including multipath, inaccurate base station coordinate, and environmental factors. Those problems are affecting in achieving the precise positioning of the rover station, which results in a less accurate position (i.e., within a few meters of accuracy), called a float solution.

Specifically, in the Himalayan, mountain, or snowy area, various errors affect positioning accuracy and reliability. Once the base station is interrupted because of errors, the recovery time is often several minutes or even a few hours. Therefore, ensuring the data continuity and reliability of the base station under challenging environments is very important, particularly in a landslide and heavy snowfall area. Recently, a drone carrying medicines and equipment service, aiming to provide medical care to the remote mountain communities and precisely measuring the altitude of mountain, including Mt. Everest, is the subject of attention in the Himalayan country, Nepal[3][4]. Similarly, many applications, including mapping, survey, and precision agriculture, are increasing day by day. Thus, the authors in this paper proposed a modality of RTK-GNSS infrastructure that could enhance the system's overall performance and reliability.

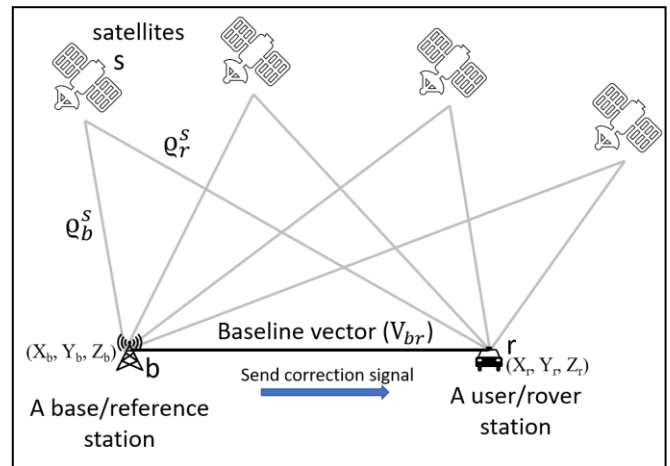


Figure 1: Working principle of differential correction

## 2 PROBLEM STATEMENT

There are various errors factor that affect the reliability of RTK-GNSS and its positioning accuracy. GNSS positioning accuracy is mainly limited by two types of errors: distance-dependent errors, and station-dependent errors[5]. The distance-dependent errors include orbital, ionospheric, troposphere delay, and station-dependent error includes multipath, antenna phase center variation, and receiver hardware biases. However, the distance-dependent errors are corrected by a base station in the RTK-GNSS system and neglected, if the base station and the rover stations are operating in the same environmental area (i.e., generally considered as 10km from the base station). Beyond this range, distance errors and the atmospheric conditions at the base and the rover station may significantly vary. Hence, these error factors cannot cancel out though differential processing. Therefore, it is ineffectual and cannot ensure the reliability of RTK-GNSS for a moving object such as autonomous vehicle aerial mapping and surveillance by Unmanned Aerial Vehicle (UAV). Thus, the outage of the base station data is a major concern, particularly for the moving rover.

Besides, no redundancy of the base station is usually available if the running base station experiences any malfunctioning or hardware bias errors. Also, the RTK-GNSS accuracy depends on other error factors, including coordinate accuracy of the base station, visible satellite information, and multipath errors. In the mountain or snowy area, where the landscape is uneven, hard frost weather, and chances of heavy snowfall, landslide, earthquake, and volcanic eruption, caused significant problems. As a bunch of snow covered the base station's antenna surface, the degradation of received GNSS signal and multipath error by snow surface as a reflector can be occurred. Thus, the differential correctional signal of a base station affects the positioning accuracy of the rover receiver. If the base station's position is displaced or changed due to landslides or other natural disasters, this inaccurate coordinate degrades the positioning accuracy of the rover receiver. Therefore, ensuring the accurate coordinate of the base station is very important before that base station. Otherwise, there is a chance of misinterpretation as well as the chance of inaccuracy. In this scenario, it is generally needed to visit

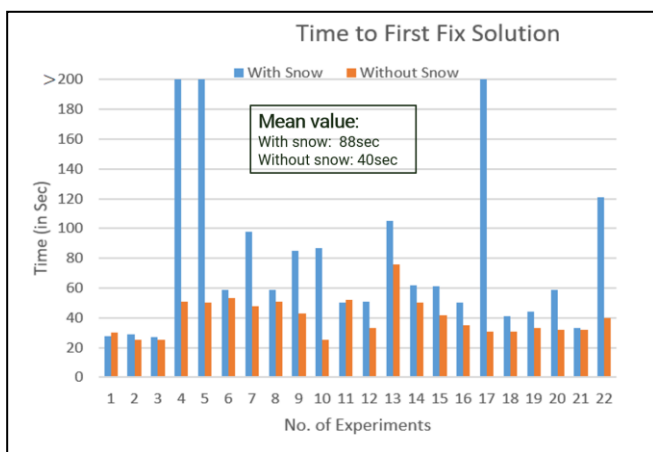


Figure 2: Histogram of converging time

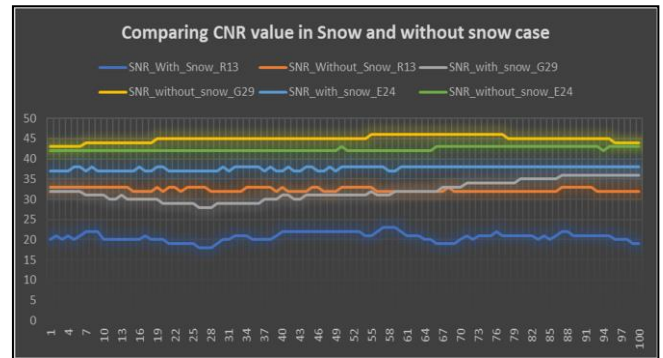


Figure 3: Observed carrier to noise data

the actual fields to repair those unchangeable error factors, but that is not a cost-effective, reliable, and appropriate solution for real-time applications. Therefore, to increase the overall performance and build a reliable infrastructure system, a base station monitoring system and the multi-base stations network with seamless handover mechanisms are needed.

## 3 BACKGROUND

In our preliminary research, we proposed a concept of cost-effective RTK-GNSS infrastructure to enhance the accuracy and feasibility of positioning solutions[6]. In that research, we mainly focused on building a base-station with sensors network and solar harvesting system. Nonetheless, the major problem was encountered while experimenting in the mountain region. Notably, due to the snow accumulation problem on the base station antenna, the precise positioning, as well as time to first fix the solution, are negatively affected. In the RTK system, once the rover receiver receives the correction message from the base station, the float solution is measured. To achieve precise positioning, the correct determination of the integer carrier phase ambiguities, called ambiguity resolution, is crucial. However, the time to get the first fix solution from the float solution is affected by snow accumulation, as shown in Figure 2. It shows that the time to first fix solution from float solution in snow existing case is significantly higher than without snow cases. Snow accumulation causes a significant problem in precise positioning while comparing carrier to noise ratio (C/N) of snow and without snow conditions. As shown in Figure 3, the carrier to noise ratio is decreased when the antenna is covered by 15cm snow height. The difference of carrier to noise ratio is around 8dB or more because of the signal diffraction inside the snow. As a result, the experiment result of more than 1m positioning error is observed. Also, distracted positioning solution and miss-fix solution was occurred in the snow accumulated antenna because of the multipath error. Likewise, previous studies have emphasized the effect of snow accumulation on the GPS antenna and verified the performance degradation and increasing of differential correction noise and multipath errors[7][8]. Besides those station-based errors, many past pieces of research point out the limitations of the current single baseline RTK while doing long-range operation with precise positioning that limits the rover receiver working area usually within 10km range from the base station[9].

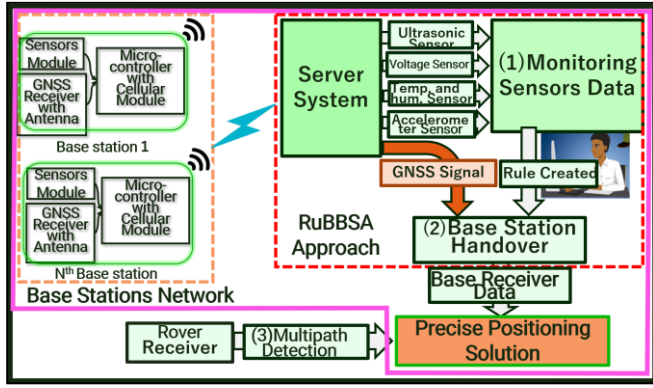


Figure 4: Complete system architecture

This is because of the spatial decorrelation of distance-dependent errors induced by the ionosphere, troposphere, and orbital errors. For long-range applications, currently, several methods can be applied, such as Network RTK, including Master-Auxiliary (MAC), Virtual Reference Station (VRS), Pseudo-Reference Station (PRS). However, these methods are challenging to implement, need active communication links, demanding control center operation, stability issues, and expensive operation costs[10][11]. Also, a seamless handover with a fix to fix solution is challenging in these systems. Generally, it happened fix to float and then fix while doing handover, which is still insufficient to the real-time moving object. Therefore, to address the station-based errors and the base station handover issues, we felt the necessities of reliable, smoothly operable, easily applicable, and cost-effective RTK-GNSS that can be used in different places and scenarios including the Himalayan and snowy regions.

## 4 SOLUTION APPROACH

This section explicitly describes the solution approach of the proposed system, mainly focused on the building of compact hardware system, sensors-based monitoring system, and continuous handover mechanism, as shown in Figure 4. The detailed explanations are as follows:

### 4.1 Building a Compact Hardware

This research is conducted by building a base station prototype module consisting of a GNSS receiver with antenna, micro-controller, and sensors network. In the rural and the Himalayan region, people live without electricity or

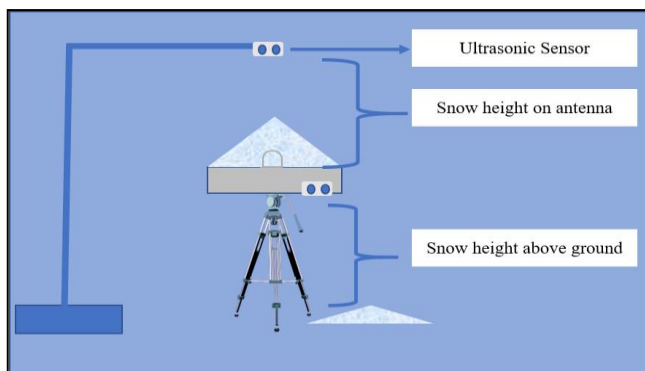


Figure 5: Snow height measurement scenario

have a severe power shortage problem. Therefore, a solar panel- an easily movable, self-sustainable, and reliable power source- is used. To process GNSS receiver and other sensors data, a micro-controller called *Wio-LTE* is used, which is a prototyping development board with LTE(4G) communication version of *Wio Tracker* (Wireless Input-output) that enables faster IoT GNSS solutions[12]. To monitor each base station's physical condition, we used various sensors modules in respective base stations. For instance, the ultrasonic sensors are used to predict the base antenna's snow height, as shown in Figure 5.

During this process, temperature fluctuation affects the sound wave an ultrasonic sensor. Thus, the temperature sensor is used in order to correct the temperature-related distortion of the measured value. Also, to precisely monitor the base station coordinate, we used a 3-axis digital accelerometer sensor that detects orientation, gesture, and motion in case of natural disasters. Besides that, to check battery voltage level and charging level, a battery state sensor is used. To make a compact hardware system, all sensors are connected in the micro-controller board that consists of a cellular modem, as shown in Figure 6. All data are sent to our server system in real-time. The base station antenna is fixed correctly at the accurate position. Similarly, to overcome the multipath errors and fully receive all visible satellite signals, the antenna is placed in an open sky environment such that all satellite signals are received as a line of sight (LOS) signal. This compact infrastructure consumes deficient power. At regular communication, the base station system needs 600mA to 2A current with a 5V power supply, making the system operation longer. Moreover, the system consists of a low-cost receiver and digital sensors, as shown in Figure 6. The total cost is around \$150. In contrast, survey-grade receivers cost around \$10,000, for instance. Therefore, our system can be comparatively cost-effective, movable, and easily installable in normal conditions and challenging weather areas.

### 4.2 Web-based Monitoring System

The first component of this research is a web-based monitoring system that is developed to supervise the state of all surrounding base stations in real-time. The base receiver data and the sensors data are collected in the centralized database through an internet connection. For a real-time application, a communication link between a base station and a user receiver is required. In choosing optimal and

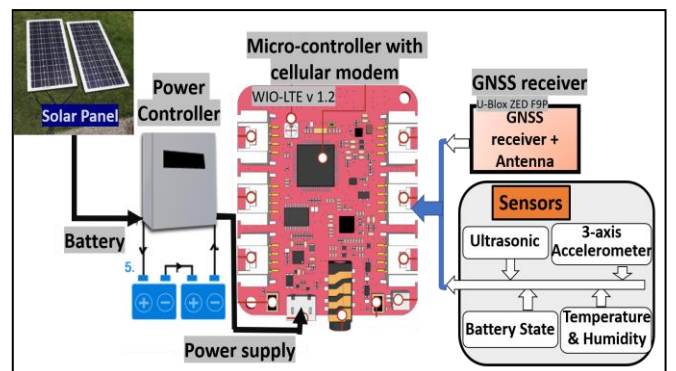


Figure 6: Block diagram of base station



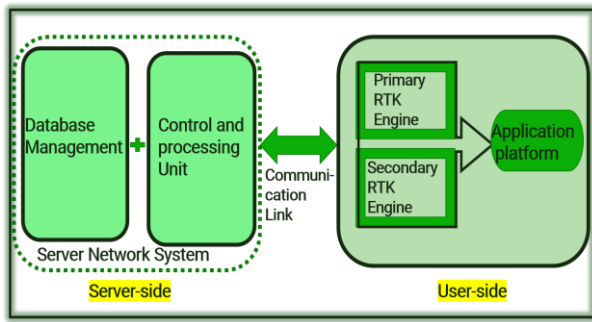


Figure 7: RuBBSA approach

effective communication methods to apply in the Himalayan region, we must consider reliability and economic factors, including operational cost, maintenance cost, and the number of computations needed by the rover and the processing center. Considering the above factors, we found that cellular modem is one of the appropriate devices in the Himalayan region for this objective. To monitor the real-time state of base stations, the server data are sent to the web-based platform to visualize the graph of all sensors data. The web-based graph is plotted on the web browser, which shows that each sensor's data is being updated with time-lapse. This process of monitoring the real state of data is adequate, especially when the base stations' knowledge is manually needed for the assignment process. More importantly, the sensors' data are used in the base assignment process to determine an optimum base station.

### 4.3 Base Station Assignment Algorithm

The second component is a processing component, an algorithm specially designed to assign the most favorable base station from the list of multiple base stations. To achieve precise positioning, we proposed a Rule-Based Base Station Assignment (hereafter, RuBBSA) algorithm. In this algorithm, the rule is based on two major factors: sensors value, and the distance between the rover receiver and the corresponding base station. The sensor's measurement is used to monitor the base station's physical condition, such as the problems, including snow accumulation, battery power supply outage, and preciseness of base station coordinates. Moreover, when the user moves out from the operating baseline area and/or conditions when the base station is unable to send differential correction information to the rover receiver, this algorithm will be activated to assign a new base station. This processing of choosing an optimum base station from multiple base stations is the main target of this algorithm. The next available optimum base station is assigned dynamically with the seamless handover from this algorithm. For this handover process, two RTK engines are simultaneously run on the rover side because it is difficult to correctly determine the integer carrier phase ambiguities in single RTK. The ambiguity resolution, which is the important factor for precise positioning, is the process of resolving the unknown cycle ambiguities of double-difference carrier phase data as integers. There are mainly three steps to determine ambiguities resolution: estimating float-valued ambiguities, finding the best integer ambiguity

set, and validating the best ambiguity set. Thus, the proposed mechanism has two RTK-engines to provide continuous and precise positioning, even in the base station assignment.

## 5 METHODOLOGY

### 5.1 Working Mechanism

This section explains the principle of the RuBBSA algorithm, and its approach explicitly, as shown in Figure 7. In this approach, there is server backend and user end. On the server-side, the server network system consists of two primary units: the database management unit, and a control unit. The database management unit is designed to manipulate and manage data. Generally, all physical sensors data, all base stations coordinate data, and real-time differential correction data of corresponding base stations are collected and then manipulated. These data are processed to the central unit called as control and processing unit, where the rule is created to specify the most favorable base station from multiple base stations. Primarily, the distance between the rover and each base station is calculated and list out all neighboring base stations on the hierarchical increasing order. The least distance is in the highest priority order. After that, physical condition of the corresponding base station, based on sensors value (i.e., snow height, antenna deflection, and battery voltage), are analyzed. If all the sensors' values are above the threshold

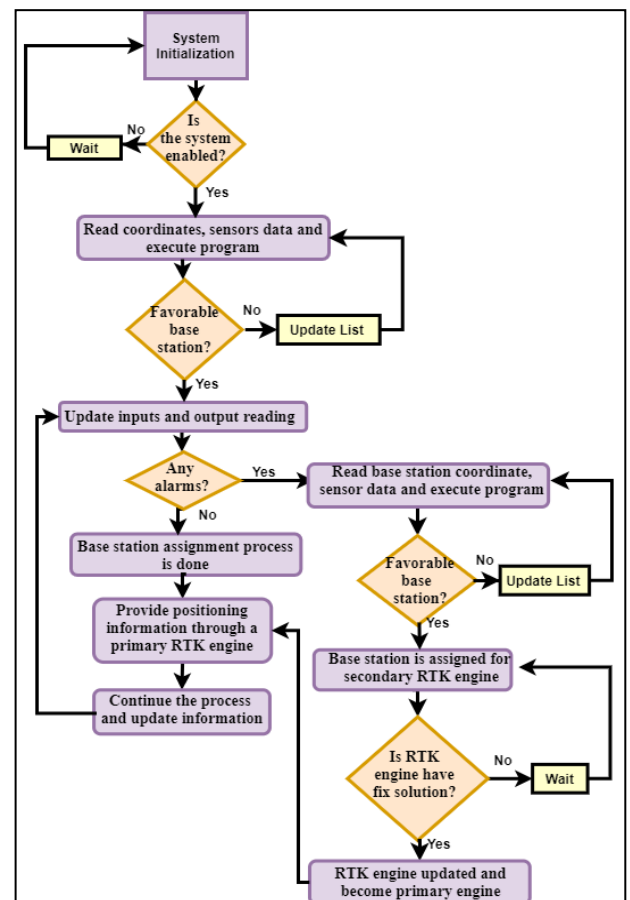


Figure 8: Flowchart of RuBBSA approach

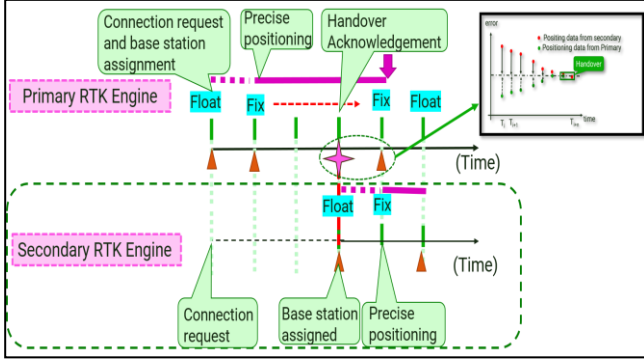


Figure 9: Base assignment process in RTK engine

value, then the optimum base station is assigned through control unit. Thus, handover is done to the proper base station. On the other hand, if the base station is out of baseline area or the sensor's value is less than the threshold value, the currently assigned base station is replaced by the next adjacent base station. In this process checking the sensors' threshold value is done before assign in RTK engine. This process will continue until the base station meets all conditions. The threshold value of sensors is based on a different scenario, such as geographical area, working hours. For instance, the threshold value was selected as 5cm after doing multiple experiments in Wakkanai, Hokkaido because the CNR value is drastically decreased and time to first fix solution is increased from that point. The flowchart of this algorithm is shown in Figure 8.

The RTK processing engine is placed on the user side, processing the differential correction signal from the base station and positioning data from the rover receiver. In this proposed system, two RTK engines, named primary and secondary RTK engines, are used to make a seamless handover operation. The primary RTK engine operates as a default RTK engine that runs until the handover is needed. The assigned base station from a control panel is linked with the primary RTK engine to provide precise positioning in the application layer.

In the case of a kinematic rover receiver, it is continuously moving from one place to another. Therefore, the positioning information of the rover receiver is updated in the control unit regularly. Similarly, the new favorable base station is assigned and linked with the secondary RTK engine, which provides a fixed solution before the primary base station falls to the float solution, as shown in Figure 9. After the handover, the secondary RTK provides a positioning solution to the application layer.

## 5.2 Determining of an Optimal Base Station

In this section, the mechanism of relative positioning is briefly introduced. As shown in Figure 1, the code pseudorange ' $\rho_b$ ' and phase pseudorange ' $\phi_b$ ' at base station 'b' to satellite 's' measured at epoch ' $t_0$ ' can be modeled by

$$\rho_b^s(t_0) = \mathbf{q}_b^s(t_0) + \Delta\mathbf{q}_b^s(t_0) + \Delta\mathbf{q}^s(t_0) + \Delta\mathbf{q}_b(t_0) \quad (1)$$

$$\lambda^s \Phi_b^s(t_0) = \mathbf{q}_b^s(t_0) + \Delta\mathbf{q}_b^s(t_0) + \Delta\mathbf{q}^s(t_0) + \Delta\mathbf{q}_b(t_0) + \lambda^s \mathbf{N}_b^s \quad (2)$$

where,  $\mathbf{q}_b^s(t_0)$ ,  $\Delta\mathbf{q}_b^s(t_0)$ ,  $\Delta\mathbf{q}^s(t_0)$ ,  $\Delta\mathbf{q}_b(t_0)$ ,  $\mathbf{N}_b^s$  are the geometric range, orbital errors, satellite-dependent errors,

receiver-dependents errors, and phase ambiguity, respectively[13]. Also ' $\lambda^s$ ' is the wavelength, defined as  $\lambda^s = c/f^s$ , where  $c$  is the speed of light and  $f^s$  is the frequency of satellite carrier. In relative positioning, the code and phase correction of the base station for same satellite at base epoch ' $t_0$ ' is calculated as:

$$\text{PRC}^s(t_0) = \mathbf{q}_b^s(t_0) - \mathbf{R}_b^s(t_0) \quad (3)$$

$$\text{PRC}^s(t_0) = \mathbf{q}_b^s(t_0) - \lambda^s \Phi_b^s(t_0) \quad (4)$$

Similarly, in the rover receiver 'r' the code pseudorange  $\rho_r$ , and phase pseudorange ' $\phi_r$ ' are calculated for the observation epoch 't' because the range and range rate correction (RRC) referring to the base epoch  $t_0$  are transmitted to the rover receiver in real time. At 'r', the pseudorange, carrier phase and the pseudorange correction (PRC) for the observation epoch 't' is modeled by

$$\rho_r^s(t) = \mathbf{q}_r^s(t) + \Delta\mathbf{q}_r^s(t) + \Delta\mathbf{q}^s(t) + \Delta\mathbf{q}_r(t) \quad (5)$$

$$\lambda^s \Phi_r^s(t) = \mathbf{q}_r^s(t) + \Delta\mathbf{q}_r^s(t) + \Delta\mathbf{q}^s(t) + \Delta\mathbf{q}_r(t) + \lambda^s \mathbf{N}_r^s \quad (6)$$

$$\text{PRC}^s(t) = \text{PRC}^s(t_0) + \text{RRC}^s(t_0)(t-t_0) \quad (7)$$

where,  $(t-t_0)$  is defined as latency. After applying the predicated pseudorange correction  $\text{PRC}^s(t)$  to the measured pseudorange of rover receiver, the satellite-dependent bias has canceled out. Also, the base and the rover receiver have highly correlated satellite-receiver specific biases in a short base-line area. Neglecting these biases, the corrected code and the phase pseudorange are calculated in the rover receiver as:

$$\rho_r^s(t)_{\text{corr}} = \mathbf{q}_r^s(t) + \Delta\mathbf{q}_{br}^s(t) \quad (8)$$

$$\lambda^s \Phi_r^s(t)_{\text{corr}} = \mathbf{q}_r^s(t) + \Delta\mathbf{q}_{br}^s(t) + \lambda^s \mathbf{N}_{br}^s \quad (9)$$

where  $\Delta\mathbf{q}_{br}^s(t) = \Delta\mathbf{q}_r^s(t) - \Delta\mathbf{q}_b^s(t)$  and  $\mathbf{N}_{br}^s = \mathbf{N}_r^s - \mathbf{N}_b^s$  are the difference of phase ambiguities[13]. To determine the coordinate of an unknown point with respect to a known point. Thus, the baseline vector between the base and the rover is calculated with corresponding position vectors  $\mathbf{X}_b$ , and  $\mathbf{X}_r$  formulated as:

$$\mathbf{X}_r = \mathbf{X}_b + \mathbf{X}_{br} \quad (10)$$

$$\mathbf{V}_{br} = \begin{bmatrix} \mathbf{X}_r - \mathbf{X}_b \\ \mathbf{Y}_r - \mathbf{Y}_b \\ \mathbf{Z}_r - \mathbf{Z}_b \end{bmatrix} = \begin{bmatrix} \Delta\mathbf{X}_{br} \\ \Delta\mathbf{Y}_{br} \\ \Delta\mathbf{Z}_{br} \end{bmatrix} \quad (11)$$

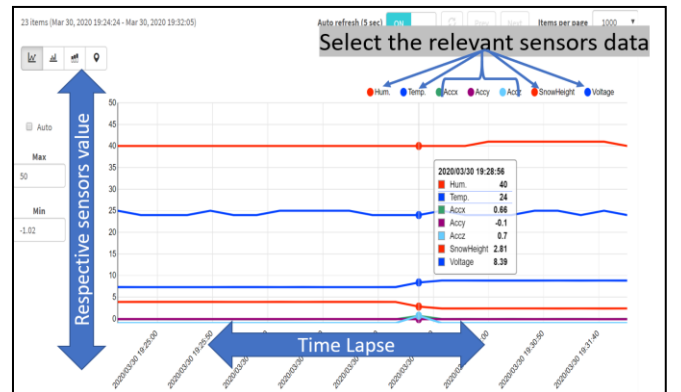


Figure 10: Graphical view of sensors data

Here, the coordinates of the base point must be accurately known to calculate the rover receiver coordinate with high precision. To find the nearest base station, the rover receiver sends its position via wireless communication to the network control center where computations are carried out. To construct the observations, the range between the base stations and the rover receiver is computed using geographic coordinate system, where the nearest continuously operating base station is selected to the user. The distance between the rover and all neighbor base station is calculated as follows:

$$\Delta D_{rb} = E * \arccos[(\sin(\text{lat}_r) * \sin(\text{lat}_b)) + \cos(\text{lat}_r) * \cos(\text{lat}_b) * \cos(\text{long}_b - \text{long}_r)] \quad (12)$$

Where, 'lat<sub>r</sub>', 'lat<sub>b</sub>', 'long<sub>r</sub>', 'long<sub>b</sub>' are the latitude of rover, latitude of base, longitude of rover and longitude of base respectively and all values are in radians. 'E' is the equatorial radius of earth and 'D' is the distance between rover and a base station.

Furthermore, the availability of base station is also measured through sensors data. For instance, ultrasonic sensors, voltage sensors, and 3-axis accelerometer sensors are used in this research. Besides these sensors, the other sensors can be used based on geographical and environmental conditions. The threshold values need to be entered at the starting time. In this rule of the assignment process, there are following three cases.

*Case I:  $S_i \geq S_{TH}$  and  $D_i \leq D_{i+1}$ ; optimum base station*

Where, 'S<sub>i</sub>' is the output of cumulative function of sensors values. The threshold value of 'S<sub>TH</sub>' is needed to set after various experiments. Also, 'D<sub>i</sub>' is the least distance between the rover, for 'i<sup>th</sup>' base station. Similarly, 'D<sub>i+1</sub>' is the distance between next adjacent base station and rover. In the above scenarios, if the base station satisfies case I, this base station is considered as an optimum base station and assign this base station till the next handover is needed.

*Case II:  $S_i \geq S_{TH}$  and  $D_i > D_{i+1}$ ; keep and hold (OK)*

If the base station satisfies case II, the base station is considered as an acceptable base station or the next potential base station. Thus, these base stations are kept and hold for the next handover. Handover may require while the rover moved far from the earlier base station and approach to the

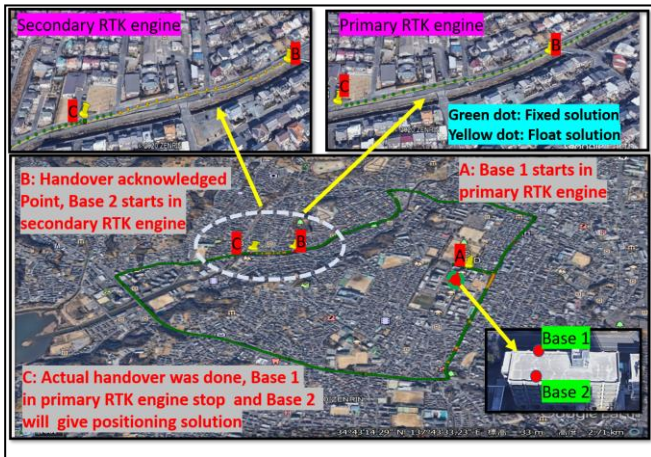


Figure 11: Experimental scenario

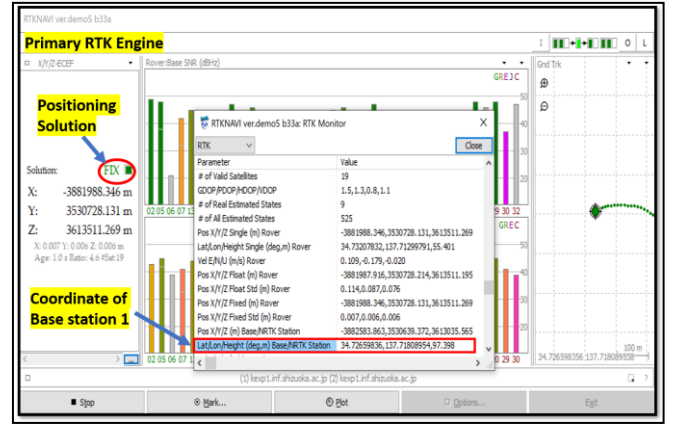


Figure 12: Positioning solution in primary RTK engine

adjacent base station. In this case, the earlier base station will be dropped off and the adjacent base station will be handover for operating base station.

*Case III:  $S_i < S_{TH}$ ; remove from the list (NG)*

If the base station satisfies case III, it is considered functionless or not a referenceable base station. Therefore, this base station is removed from the list until it satisfies cases I or II. This situation may arise due to various reasons such as due to the accumulation of snow, increase of distance between rover and base station etc.

## 6 DISCUSSION AND RESULTS

To test the performance of the proposed system, the base station was set up and tested in the practical field. Based on the reliability, and consistency of the system, two proposed components and their performance are carried. A set of task actions and obtaining results are described in the following sub-section.

### 6.1 Discussion of Monitoring System and It's Performance

Base stations equipped with digital sensor network systems were deployed throughout all base stations. We have considered three significant problems in the mountain and snowy regions that affect the base station conditions. These problems are (1) snow accumulation, (2) natural disasters, and (3) power outage problems. Therefore, to address these problems, various cost-effective and easily applicable sensors are used. To address the problem (1); ultrasonic sensors are used in the base station to measure and predict snow height. We have done our experiment in the winter of Wakkanai, Hokkaido, where there was much snow, dense fog, and a cold-weather situation. For this experiment, we have used two ultrasonic sensors. One is used for measuring snow height on the antenna, and another is used for measuring snow height in the ground surface, as shown in Figure 5.

Similarly, the 3-axis digital accelerometer is used to address the problem (2), which measures orientation, and deflection of the base antenna. This sensor is fixed along with base antenna such that small deflection or motion also measured when the base coordinate was changed by



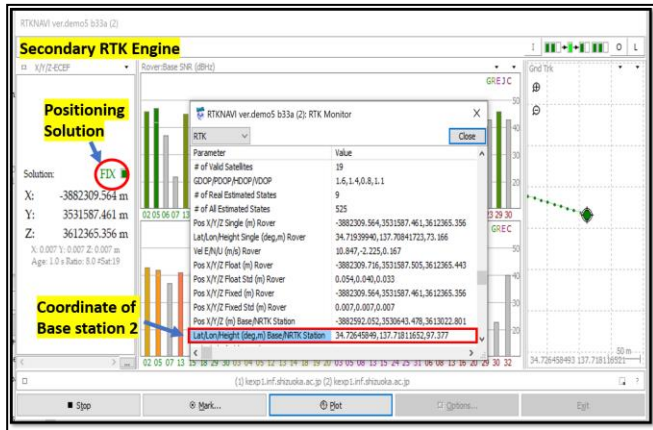


Figure 13: Positioning solution in secondary RTK engine

different factors such as some bad people intense or natural disasters. Likewise, to address the problem (3), a voltage sensor is used. The voltage level in real-time is measured and predict the operation hour of a base station. These sensors data are sent to the server network system for real-time monitoring through the web-based system and the control unit for decision making. A cost-effective and reliable approach of cellular 4G LTE connection was devised to link between a base station and a cloud server with real-time web service for users. The real-time sensor values of a base station are displayed graphically, as shown in Figure 10—this wireless data collecting, and monitoring system help to understand the physical conditions of the base station. Thus, ensuring the continuity and reliability of the base station data is possible.

## 6.2 Discussion of RuBBSA Algorithm and It's Performance

The practical use and performance evaluation of our proposed algorithm was done with static and kinematic rover receiver, as shown in Figure 11. In particular, we tested the handover process, system's functionality, and its performance concerning the RTK accuracy. To test the performance of the proposed algorithm, we have done our field test experiment in Hamamatsu Japan, where the surrounding environment contains tall buildings and dense traffic. In our kinematic experimental scenario two static base stations and one kinematic rover receiver are used. For this experiment, the base station assignment process was computed manually through a server system. At first, we have connected all base stations in the control server system where available base stations are displayed with their unique identity. As every base station consists of a cellular module as an internet provider, the unique IMSI (International Mobile Subscriber Identity) number is used for base station identity. Basically, a base station is assigned to the primary RTK engine. This selected base station starts connection to the rover receiver and get fixed solution after few seconds, as shown in Figure 12. After getting fixed solution in the RTK engine, the positioning information is ready to use by applications.

At the time of handover, the new base station is assigned through the server system to the user end. This time, the communication link is established to the secondary RTK

Table 1: RTCM message type and description

Message Type	Description
RTCM 1005	Stationary RTK reference station ARP
RTCM 1077	GPS MSM7
RTCM 1087	GLONASS MSM7
RTCM 1097	Galileo MSM7
RTCM 1127	BeiDou MSM7
RTCM 1230	GLONASS code-phase biases

engine because the first base station is still operating in primary end. In this case, the secondary RTK engine receives correctional information from a base station and provide the fix solution, as shown in Figure 13. After getting a fix solution in the secondary RTK engine, the primarily selected RTK engine went to ideal mode. The secondary RTK engine starts its positioning solution and becomes the default RTK engine. An open-source program package of RTKLIB library with a program package is used as RTK engine[14].

## 6.3 Mechanism of RTCM Corrections and RTK engine

In RTK system, the base station sends corrections to the rover via a communication link. This correction signal enables the rover receiver to compute its position relative to the base with high accuracy. Radio Technical Commission for Maritime (RTCM) is the standard format with binary data protocol for communication of GNSS correction information. The high precision receiver supports RTCM version 3 as a supported standard format. While building an RTK environment, the base receiver configuration must change to output RTCM message because we need to send standard messages and the base antenna reference point (ARP) with cm level accuracy. RTCM observation multiple signal message (MSM) will be streamed as soon as they are configured for output. The desired RTCM messages must be selected and configured for the corresponding GNSS. The list of RTCM output messages for a base operating is shown in Table 1.

In this approach, the coordinate of base station is changed, when the base station is assigned dynamically. Thus, certain time is necessary in order to compute the differential errors and to fix the ambiguities. Therefore, we must need to send the correction message as well as the base station coordinate of the corresponding base station. Besides that, certain time is needed to get the first fix solution after doing handover. This is because of the challenging process of the carrier phase ambiguity resolution. This results the positioning solution is dropped back to the float solution after handover, which is not a proper solution for the precise positioning application. Therefore, to address this dropping process during handover, we proposed an embedded technology of two RTK engine, which helps to make seamless handover.

## 6.4 Grouping of Base Stations for Network-RTK

The present study proposed a concept of network-RTK to provide a continuous and precise positioning solution for road vehicles and intelligent traffic system (ITS) applications. A fundamental aspect of network-RTK is the delivery of base station data used in the processing of the receiver's positioning and continuously provide a positioning solution with a network of reference stations. Although there are various methods used in the networking of base stations, these methods are challenging to implement, need active communication links, and demand control center operation. Therefore, we proposed the grouping of base stations so that the multiple base stations can be assigned in a single framework. Here, the available base stations are displaced with a unique number. This mechanism of grouping other than adjacent base stations can significantly reduce the controlling operation of every base station, and if the grouping is done with three controlling units, all base stations can manage properly, and seamless handover is possible for the wider base-line area too.

## 7 CONCLUSION

This paper has presented a reliable RTK-GNSS infrastructure that can enhance the overall performance and reliability of the positioning and navigation system in the mountain and snowy regions. The authors of this paper introduced a new approach of base station assignment to provide continuous and precise positioning, immediately after base station handover. Also, the web-based continuous monitoring system is practically implemented, tested, and evaluated in this paper to address major issues of RTK-GNSS. First of all, the state of the base station is measured through physical sensors and monitoring through a web-based system in order to supervise base station and find the availability. Generally, the station dependent errors are monitored through the web in real-time. The result of the experiments confirmed that the continuity and reliability of the base station data is ensured by a real-time monitoring system. Secondly, we have proposed an algorithm to assign the most favorable base station from multiple base stations based on two different factors; base-rover distance and sensors value. In this rule-based base station assignment approach, the most reliable base station is chosen to eliminate differential positioning errors. This handover strategy helps to maintain the RTK positioning at centimeter-level accuracy during the handover process. The experimental result shows that the user receiver can get the next available optimum base station and do seamless handover through the concept of two RTK engines on the user side. The authors conclude that accuracy and reliability can be achieved by using this proposed approach. The development of this infrastructure will enhance the precise positioning in challenging environments as well as different landscape areas. In this system, the rover receiver still receives both line of sight (LOS) and non-line of sight (NLOS) signal. Therefore, future research will aim to improve the positioning accuracy with the proposed method and address the issue of multipath errors for a robust RTK-GNSS.

## ACKNOWLEDGMENT

This research work was supported by JSPS KAKENHI Grant Number JP17H01731.

## REFERENCES

- [1] P. P. Das and S. Nakamura, "Analysis of GPS Single Point Positioning and Software Development," *Iject*, vol. 7, pp. 34–40, 2016.
- [2] T. Kitani, H. Hatano, M. Fujii, A. Ito, and Y. Watanabe, "A cooperative GPS/GNSS positioning method with neighboring receivers," in *Proc. of International Workshop on Informatics (IWIN2014)*, 2014, pp. 145–153.
- [3] U. Pudasaini, "Drones Optimized Therapy System (DrOTS): Use of Drones for Tuberculosis Diagnosis in Nepal," *Int. J. Hum. Heal. Sci.*, 2019.
- [4] D. Jonathan, "Trimble: Mapping Everest," *Int. Work. Meas. Height Sagarmatha GNSS Appl.*, 2017.
- [5] A. El-Mowafy, "Precise Real-Time Positioning Using Network RTK," *Glob. Navig. Satell. Syst. Signal. Theory Appl.*, October, 2012.
- [6] B. Rokaha, B. P. Gautam and T. Kitani, "Building a Reliable and Cost-Effective RTK-GNSS Infrastructure for Precise Positioning of IoT Applications," *Proc. Twelfth Intl. Conf. on Mob. Compt. and Ubinq. Net. (ICMU), Kathmandu, Nepal*.
- [7] T. Yoshihara, H. Motoyoshi, T. Sato, S. Yamaguchi, and S. Saito, "GAST-D integrity risks of snow accumulation on GBAS reference antennas and multipath effects due to snow-surface reflection," *Inst. Navig. Int. Tech. Meet. 2013, (ITM 2013)*, pp. 112–120, 2013.
- [8] J. B. Schleppe and G. Lachapelle, "GPS tracking performance under avalanche deposited snow," *Proc. Inst. Navig. - 19th Int. Tech. Meet. Satell. Div.*, vol. 5, pp. 3105–3116, September, 2006.
- [9] T. Kitani, "Bikeinformatics: An Introduction of Informatics to the Motorcycle Researches and the Development of New Generation Motorcycle-based Personal Vehicles," *Journal of Information Processing*, 2020, Vol. 28, pp. 3-15, January, 2020.
- [10] T. Maolin and F. Yanming, "Area-oriented reference station placement for network RTK," *Proc. - Int. Conf. Comput. Sci. Softw. Eng. CSSE 2008*, vol. 4, pp. 919–922, 2008.
- [11] Y. Du, G. Huang, Q. Zhang, Y. Gao, and Y. Gao, "A new asynchronous RTK method to mitigate base station observation outages," *Sensors (Switzerland)*, vol. 19, 2019.
- [12] Faire Shenzhen, "Wio LTE Cat.1," *the first maker faire in China*, 2012. [Online]. Available: [http://wiki.seeedstudio.com/Wio\\_LTE\\_Cat.1/](http://wiki.seeedstudio.com/Wio_LTE_Cat.1/).
- [13] B. Hofmann-Wellenhof, *Elementary Mathematical Models for GNSS Positioning*. 2018.
- [14] T. Takasu and A. Yasuda, "Development of the low-cost RTK-GPS receiver with an open source program package RTKLIB," *Int. Symp. GPS/GNSS*, pp. 4–6, October, 2009.

Session 4:  
Network and Security  
( Chair: Tomoki Yoshihisa )



# A Proposal of Autonomous Control of Server Relocation for Fog Computing Systems

Kouki Kamada<sup>†</sup>, Hiroshi Inamura<sup>‡</sup>, Yoshitaka Nakamura<sup>‡</sup>

<sup>†</sup>Graduate School of Systems Information Science, Future University Hakodate, Hakodate, Japan

<sup>‡</sup>School of Systems Information Science, Future University Hakodate, Hakodate, Japan  
{g2119012, inamura, y-nakamr}@fun.ac.jp

**Abstract** - Fog computing, which extends the paradigm of cloud computing to the edge of networking, has been proposed, and its research has been active. In the field of networking, research on Content Centric Networks (CCN) has been conducted. CCN have been shown to be able to handle cached content naturally within the network, reducing traffic and latency. However, in today's Internet, dynamic content with dynamic services is indispensable. A system that can handle dynamic services is desired by incorporating the way of handling computational resources in fog computing into CCN. In this paper, we propose an autonomous control of server relocation for fog computing systems for server relocation and allocating resources. In addition, Furthermore, we perform simulations on show the basic performance of the proposed system.

**Keywords:** Contents Centric Network, Fog Computing, In-Network Caching, Server Relocation

## 1 Introduction

The number of IoT devices, which is 27.4 billion as of 2017, is expected to increase to about 40 billion by 2020[1]. For these large volumes of data generated by IoT devices, processing-intensive architectures such as cloud computing do not take advantage of the processing power of the edge and the latency from the point of data generation to the remote data centers cannot be ignored. Therefore, fog computing, which extends the paradigm of cloud computing to the edge of the network, has been proposed and actively studied[2].

In the field of networking, research on CCN (Content Centric Networks) such as NDN (Named Data Networking) has been carried out instead of the conventional IP address-based architecture[3]. It has been shown that CCN can naturally handle cached content in the network by using location-independent content as an identifier, which can reduce traffic and latency.

We proposed an autonomous control of server relocation for fog computing systems[4]. In addition, we improved the autonomous control of server relocation to transfer services on the fog network where the processing capacity is heterogeneous, so that the service transfer is commensurate with the required processing capacity[5].

In this paper, to optimize end-user QoS, we control server transfers in a fog computing environment to achieve both shortening of the average response time and fairness between users. For this purpose, we set up a use cases to examine the fairness between users in uniform computer resources.

## 2 Related Works

### 2.1 Fog Computing

In fog computing, the delay time for execution is reduced by selecting and transporting the points necessary for the execution process. For example, in Wireless Sensor and Actuator Networking, simple processing can be performed at intermediate nodes, such as fog nodes, before the data collected by sensor nodes are moved to the cloud, and the intermediate nodes can reduce the delay time by giving commands to actuation instead of the cloud. There is another technique called code-offloading[6]–[9]. Code-offloading is the use of mobile on resource-constrained mobile devices. This technology aims to improve the energy efficiency and execution speed of applications. Specifically, in a mobile application, we can use the node on fog that has more computational resources to execute the code, rather than running on mobile devices. With the decision, we can save resources such as batteries in mobile devices. In fog computing, the optimal allocation of computational resources is a focus, but there has been no discussion on the optimal placement of content.

### 2.2 CCN

Jacobson et al.[3] proposed a CCN that does not use the traditional IP addressing architecture and Two types of CCN messages, Interest and Data, are used in the CCN communication. This is done by a protocol that is based on the Messages can be sent and received through the FIB( Forwarding Information Base), CS( Content Store), PIT (Pending Interest Table) to send the data back to the requester, three main data structures are used. Using these data structures, CCN exchange messages between Interest and Data. The result retains the simplicity and scalability of IP but offers much better security, delivery efficiency, and disruption tolerance. In this way, CCN put content closer to the user, which allows static contents to be disseminated. However, to treat the running system, we need to care the internal state to continue the process, it is not possible to handle in the same way to provide dynamic content and services.

There is research on cache efficiency in CCN and how to route Interest packets efficiently[10]. These studies have been discussing the treatment of static content and how efficiently distributed content can be considered as transparent, and there is no discussion on how dynamic services can be distributed and deployed on the network.

### 2.3 Necessity of Fog computing and CCN integration

In order to solve the problems we have seen from the research mentioned so far, it would be useful to consider an architecture that allows us to control the deployment of services running in the cloud and dynamically redeploy them as needed. For example, it may be possible to optimize the point of execution of services by running them closer to the user.

Since the CCN is based on the idea of replacing the current TCP / IP with the CCN, there are several discussions on static content caching schemes. However, in today's Internet, which is created by real-world TCP / IP, dynamic content with dynamic services is essential. For example, there is a web page that authenticates the user and displays information appropriate for the user. We wondered if a static content caching scheme is not enough to replace the current Internet with a CCN because of the large amount of these dynamic contents. In the study of fog computing, there is little discussion on the issue of how to place data on a fog network. Therefore, we believe that the problems in the research fields of fog computing and CCN can be mutually resolved by incorporating the way computational resources are handled in the CCN, as introduced in 2.1, into the CCN.

## 3 Challenge

We consider optimizing quality of service by allowing services running on the network to be dynamically relocated. In this paper, we focus on response time as seen by the client as quality of service. We assume that the response time is expressed as the sum of the network transmission delay and the processing time at the server. When considering the response time, we need to optimize the system in two ways: fairness of the response time among clients and minimization of the processing time. An example is shown and discussed below.

### 3.1 Use case that require fairness in delay times

There is a need for autonomous resource allocation that satisfies the fairness of delay times for participants. For example, in an Internet conferencing system, media quality for all participants may not be maintained if the server is in a single location for clients distributed in different locations on the network with different latency. Therefore, there is a need for autonomous resource allocation that satisfies the equity of delay time for participants.

### 3.2 Assumptions and Requirements for Autonomous Control of Server Relocation System

We need a system that aims at fairness in average response time and shortening of service execution time among users simultaneously. We define the service response time which is the sum of the network latency between the client and server and the processing time of the service.

In addition, the system should reduce the average response time on non-uniform computer resources. For machine learning applications, where the execution time varies greatly depending on the processing performance, the processing time of the service becomes a bottleneck due to the processing performance. Therefore, by monitoring the processing performance of each node and transferring services based on the predicted service response time, services can be transferred to nodes with appropriate processing capacity.

In this system, we assume that the client has a fixed position in the network because it communicates directly with the sensor and user. On the other hand, since servers providing services are arbitrarily located in the fog/cloud, we assume that it is possible to relocate the server by transferring the state of service execution to obtain the necessary resources to execute a process.

## 4 Proposal for Autonomous Control of Server Relocation System

In order to optimize QoS for end users, we propose the autonomous control of server relocation for fog computing to reduce both the average response time and the fairness between users, as described below. We describe each of the functions required by the system, and then we describe the basic functions of the system.

This system collects information about the computing environment of the surrounding nodes, searches for a candidate node that can minimize the service response time, and transfers the server to the selected node.

In searching for candidate nodes, it is not realistic to assume global knowledge across different computing environments such as fog and cloud. As a reasonable scope of search, we assume a routing topology of CCN interest messages when the server is regarded as a resource. It collects PCEL (Available Processing Capacity and Estimated Latency) management information for each node on the path where a message arrives and determines the server transfer based on this information.

The following sections describe the main components of the proposal, the estimation of the service processing time, the collection of PCEL information on the message arrival route, and the algorithm for selecting candidate nodes.

### Service processing time

$C$  which is the processing power of a node and  $L$  which is the amount of processing of the requested service executed by the node are represented as a two-dimensional vector to decompose the processing power of the node into CPUC and the purpose-specific unit  $T$ , respectively. The processing capacity of a fog node is represented by  $(C_C, C_T)$ , and the amount of processing required for service execution is defined as  $(L_C, L_T)$ . Based on these processing capacity and processing volume, the service processing time  $T_{est}$  is defined as follows.

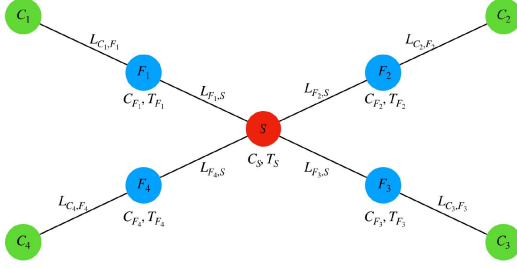


Figure 1: System Configuration Example( $S$  : Server Node,  $F$  : Fog Node,  $C$  : Client Node,  $L_{A,B}$  : Communication delay from  $A$  to  $B$ ,  $C_X$  :  $X$ 's CPU Processing Power,  $T_X$  :  $X$ 's processing of purpose-specific units)

$$T_{est}(C_C, C_T, L_C, L_T, \alpha) = \begin{cases} \frac{1}{C_C}(L_C + \frac{L_T}{\alpha}) & (C_T = 0) \\ \max(\frac{L_C}{C_C}, \frac{L_T}{C_T}) & (otherwise) \end{cases}$$

However, the CPU can perform the processing required for the purpose specific unit. The  $\alpha$  that express ratio is set to 5 for use in later evaluation.

### Collecting PCEL information on the message arrival path.

In order to treat all the nodes on the path from each client node to the server as candidates for transfer, it is necessary to collect information on the delays between the links traversed and the processing capacity of the nodes traversed. These are called the route PCEL information. In our system, PCEL information is added to the request message at the node that passes by the time the request message reaches the server node, and it is transmitted to the server.

For example, when our system is used as shown in Figure 1, the following parameters are added to the request message of  $C_1$ .

- $L_{C_1,F_1}$ , which is the communication delay of the link from  $C_1$  to fog node  $F$  when a request message from client node  $C_1$  goes through each fog node.
- $L_{F_1,S}$ , which is communication delay of the link from server node  $S$  to  $F_1$ .
- $C_{F_1}$ , which is the CPU processing power of  $F_1$
- $T_{F_1}$ , which is the processing power of the purpose-specific unit of  $F_1$

This added PCEL information can be acquired by  $S$  from the request message. Similarly,  $S$  can obtain information on the route to and from all clients from the PCEL information attached to the request messages from  $C_1$  to  $C_4$ , which are all participating client nodes.

### Candidate node selection algorithm

The server selects candidate nodes for transfer from the PCEL information appended to the request message received from the client. By using the algorithm shown in Algorithm 1. Algorithm 1 calculates the average and standard deviation of the

service response time of the participating clients based on the information that the server shown in Figure 1 can obtain from the request message. The value is the sum of the service response time multiplied by  $1 - R_{std}$  and the standard deviation multiplied by  $R_{std}$ , based on  $R_{std}$ , which specifies how much importance is placed on the fairness between clients and users. Then, we find the node whose evaluation value is at a minimum.

#### Algorithm 1 Find Candidate Node

---

**Require:**  $L_{All}$ :List of  $L$  on the Path  
**Require:**  $L_{(F_i,C_j)}$ :List of  $L$  on the Path between  $F_i$  to  $C_j$ .  
**Require:**  $F_{All}$ :List of  $F$  on the Path  
**Require:**  $C_{All}$ :Clients connected to the service  
**Require:**  $L_C$ :CPU processing capacity during service execution  
**Require:**  $L_T$ :Purpose specific unit throughput during service execution  
**Require:**  $R_{std}$ :Ratio of importance to the standard deviation  
**Require:**  $S_{F_i,C_{All}}$ :Standard deviation of service response time between  $F_i$  to  $C_{All}$   
**Require:**  $\alpha$ :Coefficient that represents the ratio when the CPU can handle the amount of processing of the target-specific unit  
**Ensure:**  $MinNode$  is candidate node.  
 $MinCost \leftarrow \infty$   
**for all**  $node$  in  $F_{All}$  **do**  
  **for all**  $client$  in  $C_{All}$  **do**  
     $Latency_{node,client} \leftarrow \sum L_{node,client} * 2$   
     $Latency_{sum} \leftarrow Latency_{sum} + Latency_{node,client}$   
  **end for**  
   $Latency_{ave} \leftarrow \frac{Latency_{sum}}{C_{All}.length}$   
   $Latency_{var} \leftarrow \frac{1}{C_{All}.length} \sum_{i=1}^{C_{All}.length} (Latency_{Ave} - Latency_{node,C_i})^2$   
   $ServiceRTT \leftarrow T_{est}(node.C_C, node.C_T, L_C, L_T, \alpha) + Latency_{ave}$   
   $R_{RTT} \leftarrow 1 - R_{std}$   
   $COST \leftarrow ServiceRTT * R_{RTT} + S_{node,C_{All}} * R_{std}$   
  **if**  $MinCost > Cost$  **then**  
     $MinCost \leftarrow Cost$   
     $MinNode \leftarrow node$   
  **end if**  
**end for**  
**return**  $MinNode$

---

### 4.1 Functions of the node

The movement of the proposed system is shown in Figure 2. This system is assumed to operate at the session layer of all participating fog nodes. It is preferable that the change in the point of service execution is done transparently to the client and server. In order to implement the collection of PCEL information and transparent addition to the request message, it is convenient to work at the session layer in the seven-layer model. Since management actions such as service transport are operated by resources below the transport layer, they must be located in the upper layer where they are visible, and



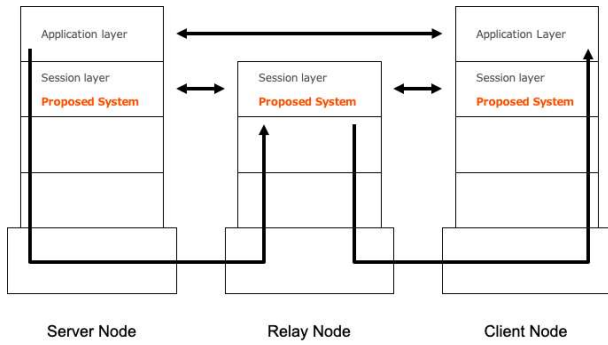


Figure 2: Autonomous Control of Server Relocation System

the session layer is lower than the application layer where clients and servers are running. By relaying communication between server nodes and client nodes at the application layer, the server relocation system at the session layer of server nodes, relay nodes and client nodes share information about resources available at each node. Based on the collected information, it realizes the selection of service execution points and resource allocation.

The proposed system consists of three types of nodes. The proposed system consists of multiple connections: a cloud node that has the contents necessary for service execution and has high processing power, a middle-class fog node that has medium processing power and can communicate with end devices with relatively low latency, and a client node that is an end device such as a smartphone that participates in the server. Based on the client-server communication model, cloud nodes and fog nodes play the role of servers, and leaf nodes play the role of clients. The server monitors the communication status of participating clients and decides whether it should autonomously play the role of the server or delegate the role of the server to other fog nodes based on the communication status. The delegated fog node takes over the role of the server. In this way, we try to optimize the server relocation of the server's role. This is an attempt to reduce the service response time.

There are Each fog node has an in-network resource monitoring function, a candidate selection function and a service transfer function. In this section, each function is explained. The autonomous control of server relocation system at each node operates at the session layer to superimpose management information on the communication messages between the server and the client to realize the resource monitoring function in the network. At the same time, it constantly monitors changes in the resources in the network, and if a change is observed, it executes the candidate selection function and, if it is judged to be necessary, it executes the service transfer function to the selected node to optimize the server relocation.

#### 4.1.1 Overall Flow to Optimize Server Placement

We summarize the optimization process explained so far. The client sends a request to the server. The server stores information associated with the request by means of an in-network monitoring function. Using the candidate selection function, we select candidate nodes from the accumulated information.

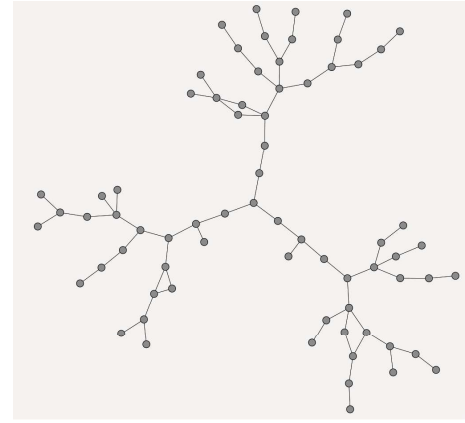


Figure 3: The network topology used in the experiment

After a candidate node is selected, it sends a message to the candidate node that it will transfer the service. A fog node that receives a message to transport a service confirms that no other service is established at its own node and starts collecting data for service execution, while at the same time sending a message to the node from which the service is being transported to inform it that the service is being prepared. When a fog node completes its data collection, it starts the service and sends a message to the source node telling it that it is ready. The server node that receives the ready message announces the new server to its current clients. The client receives information about the new server and changes the destination of the request to the new server. A client that joins from the middle of the process first sends a request to the original server node, receives information on the current server, and joins the service based on that information.

## 5 Evaluation

We defined three use cases to show three aspects: fairness between users on uniform computer resources, reduction of average response time on heterogeneous computer resources, and fairness and reduction of average response time between users on heterogeneous computer resources.

### 5.1 Simulation Environment

For the network topology, we used the topology generated by BRITE[11], which is a topology generator as shown in Figure 3. A county of three AS-equivalent nodes was prepared, with about 20 Fog nodes in each AS, and in each simulation, one AS was treated as a cloud environment and two AS were treated as a fog network with clients connected to it. Table 1 shows the parameters used in the simulation. The amount of content cache space owned by each node is also determined by the This was done assuming that the system has enough space to cache all the necessary data.

### 5.2 Internet Conference

this use case assumes a multi-point Internet conferencing system to show fairness among users with uniform computer resources. In the Internet conferencing system, the server mixes media data received from all connected terminals and

Table 1: Simulation Parameters

Parameters	Value
Cache Algorithm	LRU
Data Rate	10Mbps
Delay	1ms
Simulation time	100s
Server's $C_C$	100.0
Server's $C_T$	100.0
Dedicated Unit's $C_C$	20.0
Dedicated Unit's $C_T$	50.0
Regular Unit's $C_C$	20.0
Regular Unit's $C_T$	0

distributes them as a single stream to all terminals. The goal is to keep media quality fair in situations where geographically distributed participants connect to the system. Simulate the behavior of a server moving to the optimal location for clients distributed in different locations on the network with different latency times.

### 5.2.1 Simulation Scenario

Multiple meetings were defined and the participants of each meeting were placed in the same AS, and the servers of all the meetings were placed together in a different AS than the AS in which the clients were participating. The processing capacity for conducting the conference and the processing capacity of each node were assumed to be constant. The results were compared with the case in which no transfer was performed.

## 6 Results and discussion

In this use case experiment, we achieved fairness between users on a uniform computational resource. The experimental results are shown in Figure 6 and Figure 7. Figure 6 shows the change in service response time when the proposed system is not used, and Figure 7 plots the change in service response time against time when the proposed system is used. The users of the proposed system are gradually transferred to the one with less network latency.

At the timing of the start of the simulation, the two conference streams are shown in Figure 4. It is sent from different AS to a single AS, and the network traffic is aggregated. In the network after the transfer, the servers are transferred within each AS, as shown in Figure 5, and the two conference avoids the aggregation of meeting traffic.

In Figure 5, the fairness between users depends on which node on each communication path the server will be transferred to. It is possible to achieve the fairness required by each application by adjusting the  $R_{std}$  used in the algorithm 1 for destination determination. Figure 8 shows the trend of the mean and standard deviation of the service response time for a single conference among the results of the selecting candidate nodes for transfer, where the value of  $R_{std}$  is set to 0 and only the mean of the service response time is important. Figure 9 sets the value of  $R_{std}$  as 0.7 and the node selection with a weighted standard deviation of 70% of the response time, showed the average service response time for the same meet-

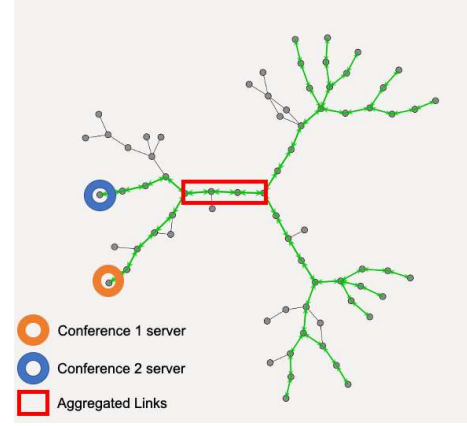


Figure 4: Network status before the transfer begins.

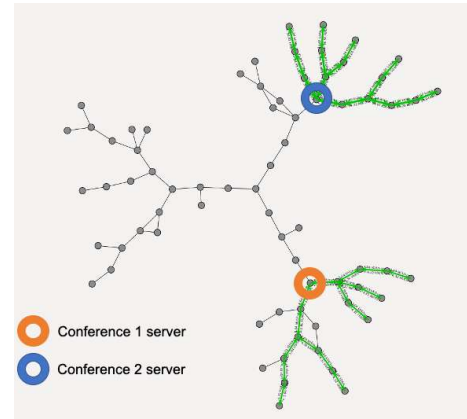


Figure 5: Network status after transfer

ing as in figure 8. Comparing the two figures, we can confirm that the result of figure 9, weighted at 70%, is fairer (i.e. smaller deviation) than the result of figure 9, which shows the fairness of the transfer between users. We can confirm that the fairness of the transfer between users is maintained. In this way, we achieved fairness among users with uniform computer resources by performing transfers to shorten the service response time and adjusting the parameters to meet the requirements of the application.

## 7 Conclusion

Fog computing, which extends the cloud computing paradigm to the edge of the network, has been proposed and is being actively researched. In the field of networking, there is research on CCN that use location-independent content as identifiers instead of the traditional IP address-based architecture. So far, we have proposed a system that aims at fairness in response time and shortening of service execution time between users, respectively. Therefore, in this study, we proposed the autonomous control of server relocation for fog computing systems to optimize QoS for end users, which achieves both shortening the average response time and fairness between users. To this end, the system achieves inter-user fairness on uniform computer resources, reduction of average response time on non-uniform computer resources, and We tested the fairness between users by setting up use cases .

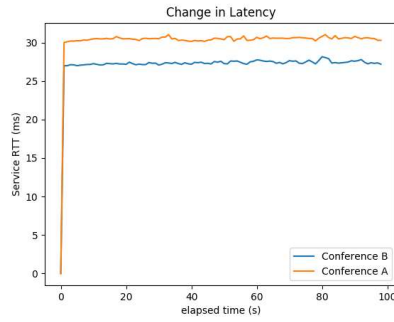


Figure 6: Changes in service response time if the proposed system was not used.

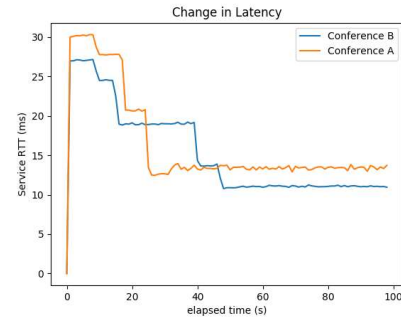


Figure 7: Changes in service response time when using the proposed system

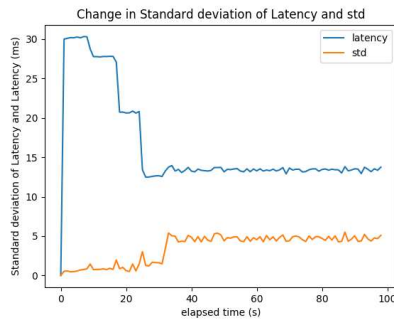


Figure 8: Mean and standard deviation of service response time for conference A when  $R_{std}$  is set to 0%.

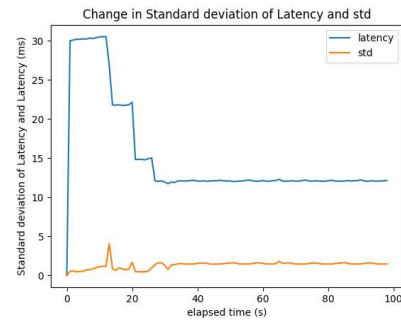


Figure 9: Mean and standard deviation of service response time for conference A when  $R_{std}$  is set to 70%.

## REFERENCES

- [1] Ministry of Internal Affairs and Communications, Japan: *The 2018 White Paper on Information and Communications in Japan*, Japanese Government (2018).
- [2] Bonomi, F., Milito, R., Zhu, J. and Addepalli, S.: Fog Computing and Its Role in the Internet of Things, *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, MCC '12, pp. 13–16 (2012).
- [3] Jacobson, V., Smetters, D. K., Thornton, J. D. et al.: Networking Named Content, *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '09, pp. 1–12 (2009).
- [4] Kamada, K., Inamura, H. and Nakamura, Y.: A Proposal of Autonomous Control of Server Relocation for Fog Computing Systems(In Japanese), *IPSI SIG Technical Report*, Vol. 2018-MBL-89, No. 1, pp. 1–5 (2018).
- [5] Kamada, K., Inamura, H. and Nakamura, Y.: Autonomous Control Scheme of Server Relocation for Non-Uniform Computing Capacity in Fog Networks(In Japanese), *DICOMO2019*, Vol. 2019, pp. 1204–1211 (2019).
- [6] Cuervo, E., Balasubramanian, A., Cho, D.-k. et al.: MAUI: making smartphones last longer with code offload, ACM Press, p. 49 (2010).
- [7] Chun, B.-G., Ihm, S., Maniatis, P. et al.: CloneCloud: elastic execution between mobile device and cloud, ACM Press, p. 301 (2011).
- [8] Kosta, S., Aucinas, A., Hui, P. et al.: ThinkAir: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading, *2012 Proceedings IEEE INFOCOM*, pp. 945–953 (2012).
- [9] Berg, F., Dürr, F. and Rothermel, K.: Increasing the Efficiency of Code Offloading in N-tier Environments with Code Bubbling, *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MOBIQUITOUS 2016, ACM, pp. 170–179 (2019).
- [10] Shanbhag, S., Schwan, N., Rimal, I. and Varvello, M.: SoCCeR: services over content-centric routing, *Proceedings of the ACM SIGCOMM workshop on Information-centric networking - ICN '11*, ACM Press, p. 62 (2011).
- [11] Medina, A., Lakhina, A., Matta, I. and Byers, J.: BRIT: An approach to universal topology generation, *MASCOTS 2001, Proceedings Ninth International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, IEEE, pp. 346–353 (2001).

# On Improving Efficiency of CSMA/CA with RSSI-based Control-frame Detection

Yoshito Umezawa<sup>†</sup> and Takuya Yoshihiro<sup>‡</sup>

<sup>†</sup>Graduate School of Systems Engineering, Wakayama University, Japan

<sup>‡</sup>Faculty of Systems Engineering, Wakayama University, Japan  
{s206036, tac}@wakayama-u.ac.jp

**Abstract** - CSMA/CA has been known as a representative media access control method since the dawn of wireless communication. Even now, it is widely used in the world. For example, it is well known to be adopted by IEEE 802.11, which is one of the most popular communication standards. CSMA/CA has long been known to have a problem that significantly degrades communication performance, called the hidden-terminal problem or the exposed-terminal problem. These problems have been tackled by many researchers over the years, and there is a huge amount of research. However, a fundamental solution to these problems has not yet been proposed. For this reason, even now, the communication performance is still significantly deteriorating when many terminals gather. In this paper, we propose a control frame multiplexing technique to detect CTS frame and ACK frame with high accuracy without demodulation by monitoring the received signal strength, i.e., RSSI (Received Signal Strength Indication), even when a node is receiving signals from neighboring nodes, which would normally be busy. The proposed technique enables simultaneous data communication, which solves the problem of exposed terminals and greatly improves the communication efficiency in CSMA/CA.

**Keywords:** CSMA/CA, RTS/CTS, RSSI, exposed-terminal problem.

## 1 INTRODUCTION

IEEE 802.11, which was standardized in 1997, is one of the most popular wireless communication standards even today. In this IEEE 802.11, a medium access control method called CSMA/CA is adopted. In CSMA/CA, a node detects if other nodes are transmitting before starting transmission. When no other node is transmitting, it starts communication after waiting a random backoff time. If a node detects that another node is transmitting, it waits for a while and after the node finishes the transmission, it starts its own communication. However, CSMA/CA has problems called the hidden-terminal problem and the exposed-terminal problem which significantly deteriorates communication performance. The hidden-terminal problem is a problem in which, when nodes that cannot detect the transmit radio waves of each other simultaneously, a collision occurs at the receiving node. The exposed-terminal problem is a problem in which transmission is actually possible, but transmission is unreasonably suppressed when there is transmission around there even if it does not prevent the transmission.

Many researchers have been tackling on these two problems for many years, and so there is a vast amount of studies.

However, no essential solution has been proposed yet.

The purpose of this study is to eliminate the exposed terminal problem in wireless communication using CSMA/CA. We propose a control frame multiplexing technique that can detect the arrival of CTS and ACK frame with high accuracy by monitoring RSSI and transmit data even when a node is in BUSY state under the original CSMA.

This paper consists of five sections. In Section 2 describes related work. Section 3 describes the method proposed in this study. Section 4 describes the performance evaluation. Section 5 describes a summary of this study.

## 2 RELATED WORKS

Improvement in CSMA/CA has been undertaken by many researchers for many years, and so there is a vast amount of studies. Bharghavan et al. proposed a method called RTS/CTS to solve the hidden terminal problem that occurs in CSMA [1]. This method is also adopted in IEEE 802.11 standard. However, it is known that the performance degradation due to the exposed terminal problem is significant, and that frame loss due to radio interference from a distance occurs frequently especially during high-speed communication, so it does not work well as a countermeasure for the hidden terminal problem [2] [3]. As a result, RTS/CTS is rarely used in practice.

Recently, methods have been proposed to improve communication efficiency by using techniques in the physical layer. In wireless communication, a technique called SIC (Self Interference Cancellation) has been proposed in which a node has a two NICs (Network Interface Cards) for transmission and reception, respectively, and cancels the transmitted signal at the receiver to perform transmission and reception at the same time. This technique is known as full-duplex wireless communication, which has been actively studied [4] [5]. In addition, there is a technique called NOMA (Non-Orthogonal Multiple Access). When a node receives a strong signal and a weak signal at the same time, by demodulating the strong signal first, and by estimating its original signal to be removed from the received signal, the node can demodulate the weak signals [6]. Although these techniques are drawing attention as ones that significantly improve wireless communication capacity, only a few techniques for utilizing them in the MAC layer have been proposed. Therefore, it is doubtful whether or not it will contribute to solve the hidden terminal problem, which is an essential problem in CSMA/CA.

J.J.Garcia-Luna-Aceves proposed CRMA as a MAC protocol using SIC technology [7]. In addition, he proposed to use busy tone and pilot signals to realize a complete MAC protocol that does not cause both hidden and exposed termi-

nal problems in wireless communication [8]-[10]. However, all of them are analyzed only theoretically, and so the performance in an actual wireless environment is unknown.

The purpose of this paper is to realize a MAC protocol that does not cause both hidden terminal problems and exposed terminal problems. The proposed method differs from the conventional method in that CTS and ACK frames can be received without demodulation. Since they are not demodulated, CTS and ACK frames can be received even if the signal from another node is being received only if the SN ratio is larger than about 3 dB. Therefore, it has the potential to greatly improve the flexibility of the MAC protocol compared to the conventional method.

### 3 PROPOSED METHOD

#### 3.1 Overview

In the proposed method, nodes constantly monitor RSSI (Received Signal Strength Indication) during wireless communication using RTS/CTS. Even if the RSSI level reaches the threshold to transit to BUSY state or the NAV state, in the original CSMA if the certain condition is satisfied, it remains in idle state and transmitting RTS/CTS frames to start DATA transmission is allowed. In addition, after data frames are sent, CTS or ACK as the response is detected without demodulation by observing only the rise of RSSI at the timing when CTS or ACK is returned. As a result, we can achieve simultaneous communication of data frames and, solves the exposed terminal problem, and improves communication performance.

An example of the operation of the proposed method is described in Figure 1 and 2. Figure 1 shows the arrangement of nodes and the communication flow, and Fig. 2 shows a MAC operation of each node. First, RTS/CTS handshake is performed between node  $s_1$  and node  $r_1$ . After that,  $s_1$  starts transmitting DATA frame when CTS reception from  $r_1$  is completed.  $s_2$  that receives RTS and the data frame of  $s_1$  does not transit to NAV or BUSY state because RSSI of RTS and DATA frame is below the predefined threshold.  $s_2$  sends RTS to node  $r_2$  after waiting the backoff time.  $r_2$  returns CTS after receiving RTS from  $s_2$ . When CTS is returned from  $r_2$ ,  $s_2$  is detecting DATA frame transmitted by  $s_1$ . However,  $s_2$  detects the rise in RSSI at the timing when CTS will be returned. As a result, it is confirmed that CTS has been returned from  $r_2$  without demodulation, and  $s_2$  starts DATA frame transmission.

When  $r_1$  finished receiving DATA frame from  $s_1$ , it returns an ACK frame. When ACK is returned from  $r_1$ ,  $s_1$  is detecting DATA frame from  $s_2$ , but the rise in RSSI is observed at the timing when ACK will be returned. As a result,  $s_1$  confirmed that ACK has been returned from  $r_1$  without demodulation, and  $s_1$  completes transmission. On the other hand,  $r_2$  received DATA frame from  $s_2$ , returns ACK, and the transmission of  $s_2$  is completed. If DATA frame of  $s_2$  finishes its transmission before DATA frame of  $s_1$ , ACK from  $r_2$  interferes with DATA frame of  $s_1$ . However,  $s_2$  judges that ACK has arrived because of the rise of RSSI, and completes the communication. In this way, the communication of  $s_2$  is not

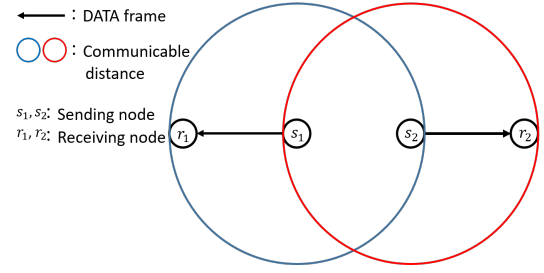


Figure 1: Layout of the operation example

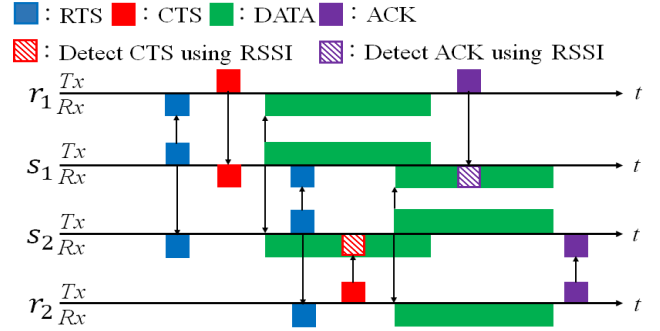


Figure 2: Operation example of each nodes

suppressed by RTS, and the simultaneous communication of DATA frames from  $s_1$  and  $s_2$  is possible, where the exposed-terminal problem is solved.

#### 3.2 RSSI-based CTS and ACK detection

In this paper, even if frames are sensed, nodes do not transit to BUSY state when a certain condition is met, and the nodes start RTS/CTS handshake on their back-off expiration. This aims to eliminate the influence of the exposed-terminal problem and improve communication throughput. However, if a node sends a RTS or DATA frame when some other nodes are transmitting DATA frames, the returned CTS or ACK frames could collide with the DATA frame, resulting in retransmission of those frames.

Our approach to prevent this is to confirm the arrivals of CTS or ACK by observing only RSSI even if CTS or ACK is not high enough to demodulate it. At this time, the timing at which CTS or ACK frames will be returned depends on the fixed-length SIFS and the frame transmission rate, meaning that the arrival time can be easily expected. Therefore, if the rise in RSSI is observed at the timing when CTS or ACK is expected, it will be the reception of CTS or ACK. Even if demodulating CTS and ACK frames is impossible, the node can confirm that CTS and ACK frames have arrived.

This is explained in Fig. 3. Node B has data for node C, but at this moment B is receiving the data frame from neighboring node A. B holds the average radio signal strength  $S_A$ [dBm] of A. At this time, if the radio signal strength being observed is less than or equal to  $S_A + T$ [dBm], where T is a predefined threshold value, B starts data transmission with RTS. If node



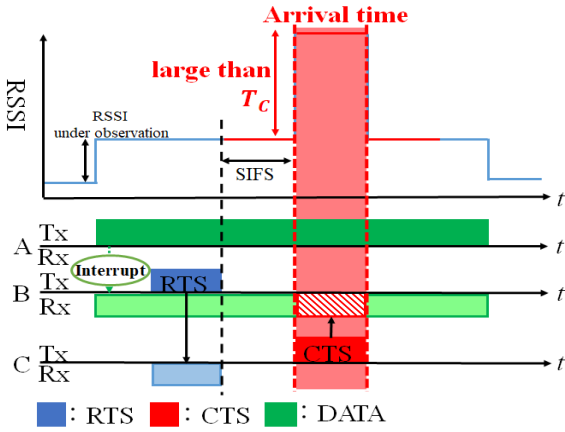


Figure 3: CTS detection using RSSI

C receives RTS of node B normally, node C returns CTS after SIFS interval. However, this time, B cannot decode CTS from C due to the interference from the data frame from node A. Therefore, node B compares the RSSI level of the estimated period for CTS arrival with that around the estimated arrival period, and if this difference is more than the threshold  $T_c$ , it judge that CTS has arrived and processes CTS. In addition, when the data transmitting nodes, i.e., A and B, are close to each other, RSSI level of DATA frames received by each other become high, and RSSI level of CTS and ACK becomes relatively small, consequently they would fail confirming CTS and ACK with high probability. Therefore, if the radio signal strength observed by B is  $S_A + T$  or more, B transits to BUSY state as in the conventional CSMA/CA, and waits for A finishing its transmission.

By detecting CTS or ACK using RSSI, communication is not suppressed by RTS or DATA frames transmitted from other nodes. As shown above, even under some interference, RTS/CTS handshake, data transmission and ACK transmission are possible, and consequently the exposed-terminal problem is resolved.

### 3.3 Proposed MAC protocol

Since the proposed method detects CTS and ACK through RSSI described in Section 3.2, a part of the conditions for state transition differs between the conventional CSMA/CA and the proposed method. Figure 4 shows the state transition diagram of the proposed method.

The difference between CSMA/CA and the proposed protocol is two folds : (1) the behavior when receiving RTS that does not destine itself, and (2) the behavior in face of carrier sensing. (1): In CSMA/CA, when RTS or CTS that does not destine itself is received in any of BACKOFF, DATA\_WAIT, CTS\_WAIT, and ACK\_WAIT states, it transits to NAV state. In contrast, in the proposed method, the transmitting node in the BACKOFF state continues to stay BACKOFF state when it receives RTS to the node if the RSSI is below the threshold. If the received RSSI is above the threshold, it transits to BUSY state, which is the same as the normal CSMA/CA, because the returned CTS would fail to be detected with high

probability. In carrier sensing, when the RSSI level of the received signal is above the threshold, it transits to the busy state.

(2): In CSMA/CA, DATA frame transmission starts only when CTS is received, and after the transmission ends, the node enters ACK\_WAIT state. However, in the proposed protocol, when CTS is received in CTS\_WAIT state, or when CTS is detected by RSSI as described in Section 3.2, DATA frame transmission starts and the node transits to ACK\_WAIT state. If the received RSSI level is above the threshold, it transits to BUSY state. When CTS is not received, the node transits to the BACKOFF state. In the proposed protocol, transition to the BACKOFF state occurs either when ACK is received in ACK\_WAIT state, or when ACK is detected by RSSI, or when a timeout occurs without receiving ACK.

Next, the common parts between the proposed protocol and CSMA/CA are described. First, the operation is the same in BUSY state and NAV state. When in BUSY state, if there is no radio in the communication channel, nodes transits to the backoff state. In NAV state, it transits to the backoff state when NAV period ends. In CSMA/CA and the proposed protocol, when a CTS for the node is received in the BACKOFF state, the state transits to NAV state. In the backoff state, the node waits before transmission until the random backoff time expires, and when the backoff expires, RTS is sent and transits CTS\_WAIT state.

## 4 EVALUATION

### 4.1 Evaluation Method

We compare the performance of the proposed method and the existing method using the network simulator Scenargie ver. 2.1. The existing methods to compare are CSMA/CA (with RTS/CTS) and CSMA (without RTS/CTS). In the experiment, we evaluate whether the communication performance is improved by solving the exposed terminal problem by the proposed method compared with the conventional method. Therefore, we focus on and evaluate the frame delivery rate and average throughput. The frame delivery rate represents the rate of the received data frames out of the number of transmitted frames. The average throughput represents the amount of data sent/received per unit time. The communication speed of the entire network is measured by this measure.

Figure 5 shows the simulation scenario. We prepare two Access Points (AP), and one AP communicates with four Clients (C). The location of each AP and C is set so that both the hidden-terminal problem and the exposed-terminal problem occur. The distance between two APs, and the distance between C and the neighboring AP are both 250[m]. The simulation time is 120[s], and the communication flow generation time is 10 to 110[s]. The communication standard used IEEE802.11g, which is a commonly used wireless communication method, and the communication speed of all nodes is 6[Mbps]. The communication flow is CBR(Constant Bit Rate) and the frame size is 1000[Byte]. Then, the simulation was performed with variation of transmission rate per flow from 50 to 900[kbps] with 50[kbps] interval. Table 1 summarizes the conditions common to both the proposed method and

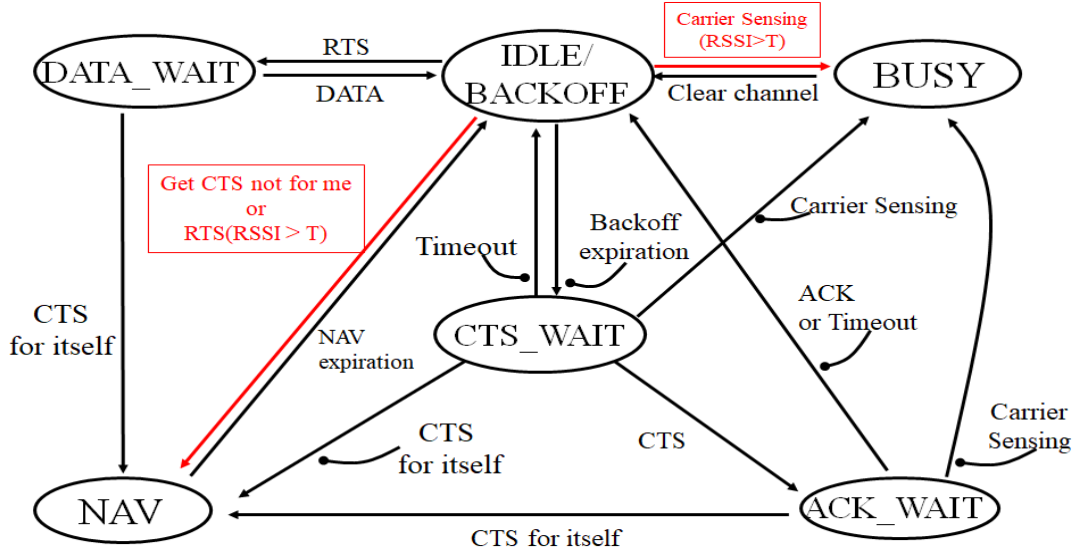


Figure 4: State transition diagram of proposed protocol

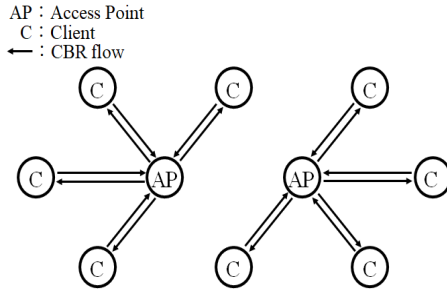


Figure 5: Node position

Table 1: Common condition

Parameter	Value
Threshold	3[dBm]
Simulation time	120[秒]
Number of nodes	10
Distance	250[m]
Flow type	CBR
Number of flows	16
Bit rate	50~900[kbps]
Occurrence time	100[s]
Frame size	1000[Byte]
Communication standard	IEEE802.11g
Communication speed	6[Mbps]
Transmission power	10[dBm]

the existing method. We set the detection threshold for CTS and ACK to 3 [dB] by t-test. This value is computed from a simple statistical calculation; we found we can identify CTS and ACK signal with 99[%] in probability if two randomly distributed signals have 3[dB] difference in the average signal strength.

## 4.2 Results

Figure 6 and 7 show the evaluation results of the simulation. Figure 6 shows the average throughput of the proposed method and the existing method. Figure 7 shows the average delivery rate. The horizontal axis represents the transmission rate of one communication flow from 50 to 900[kbps], and the vertical axis represents the average throughput [kbps] in Fig. 6 and the average delivery rate [%] in Fig. 7. In Fig. 6, we see that the performance of the conventional method and the proposed method are the same in throughput up to 400 kbps in the transmission rate. However, over 400 kbps, the proposed method exceeds the existing method and the difference reaches about 1000 kbps at the transmission rate of 700 kbps. In Fig. 6, the performance in delivery rate is the same between the conventional method and the proposed method up to 400

kbps in the transmission rate. Over 400kbps, the proposed method exceeded the delivery rate at all transmission rates. This is because the exposed-terminal problem was solved by detecting CTS/ACK using RSSI, and the communication opportunity was not lost by solving the exposed-terminal problem, and consequently that simultaneous data communication was performed. Specifically, the DATA frames transmitted from the AP on the left side reach the AP on the right side, but the AP on the right side transmits RTS or DATA frame without suppressing the transmission, so the exposed terminal problem is solved. Also, when CTS/ACK arrives at the right AP while the data frame of the left AP is arriving, the arrival of CTS or ACK is detected by RSSI, so the hidden terminal problem can be solved. From the results of the evaluation, it is clear that the proposed method solves both the hidden-terminal problem and the exposed-terminal problem to some extent. However, even with low transmission rates, the delivery rate is about 63[%], so it can be considered that it has not been completely resolved. From the above, it is clear



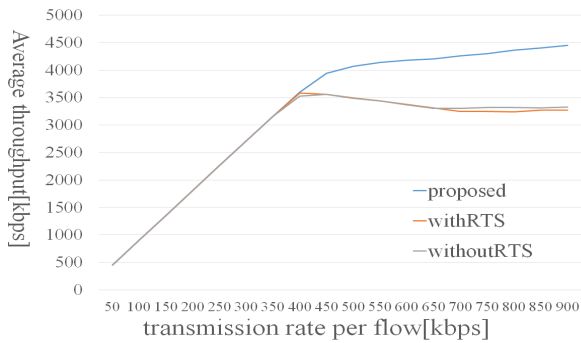


Figure 6: Average throughput results

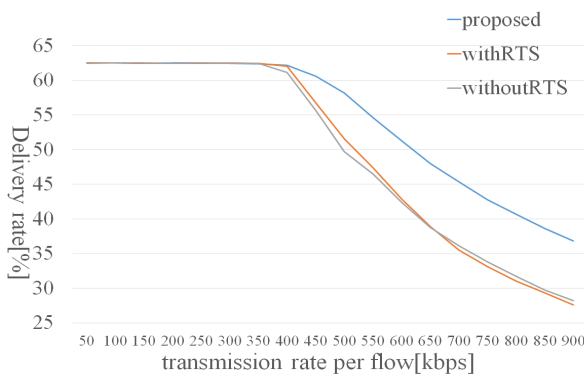


Figure 7: Average delivery rate result

that the proposed method improves the communication performance by eliminating the hidden-terminal problem and the exposed-terminal problem to some extent, as compared with the existing methods. In the proposed method, the capacity of the entire network is increased, and both the throughput and the delivery rate are improved by collision avoidance and simultaneous data communication.

## 5 CONCLUSION

In this paper, we proposed a method to solve the exposed terminal problem in wireless communication using CSMA/CA. As an evaluation, we compared the proposed method with the existing method using the network simulator Scenargie ver. 2.1. The evaluation results reveal that the proposed method improves the performance by simultaneous transmission of data, eliminating the hidden-terminal problem and the exposed-terminal problem to some extent. However, since it has not been completely resolved, we will analyze in detail the hidden-terminal problem and the exposed-terminal problem that could not be solved in the future. In addition, it is considered that the proposed method is open to discussion on the effects of false detection of CTS/ACK and of the fading effect in a certain environment. Therefore, we think that it is necessary to analyze the physical layer behavior by using MATLAB or real machine implementation. In the future, we will investigate the reality in a real environment by actual experiments using

software defined radio such as USRP.

Recently, techniques for improving wireless communication efficiency by a physical layer has been proposed, and a full-duplex wireless communication technology for simultaneously performing transmission and reception has been proposed. we would also try to apply this study to not only the current half-duplex wireless communication but also the full-duplex wireless communication.

## Acknowledgment

A part of this work is supported by KAKENHI(19K22845)

## REFERENCES

- [1] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A media access protocol for wireless LAN's," in Proc. ACM SIGCOMM '94, pp. 212–225, 1994.
- [2] J.L. Sobrinho, R. de Haan, J.M. Brazio, "Why RTS-CTS Is Not Your Ideal Wireless LAN Multiple Access Protocol," In Proc WCNS '05, 2005.
- [3] K. Xu, M. Gerla, and S. Bae, "Effectiveness of RTS/CTS Handshake in IEEE 802.11 Based Ad Hoc Networks," Ad Hoc Networks, Vol.1 Issue.1, pp.107–123, 2003.
- [4] M. Jainy et al., "Practical, Real-Time, Full Duplex Wireless," In Proc. ACM MobiCom '11, 2011.
- [5] D. Kim, H. Lee, and D. Hong A Survey of In-Band Full-Duplex Transmission: From the Perspective of PHY and MAC Layers Perspective of PHY and MAC Layers, IEEE Communications Surveys & Tutorials 17(4), 2017–2046, 2015.
- [6] Z. Ding, X. Lei, G.K. Karagiannidis, R. Schober, J. Yuan, V. Bhargava, A Survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends, IEEE Journal on Selected Areas in Communications, 35(10), pp.2181–2195, 2017.
- [7] J.J. Garcia-Luna-Aceves, Carrier Resolution Multiple Access, In Proc. PE-WASUN '17, 2017.
- [8] J.J. Garcia-Luna-Aceves, "CTMA: A More Efficient Channel Access Method for Networks with Hidden Terminals," PE-WASUN '17, pp 9–16, 2017.
- [9] J.J. Garcia-Luna-Aceves, Busy-Tone Multiple Access with Collision Avoidance and Detection for Ad-Hoc Networks, In Proc. of IEEE ICNC2019, 2019.
- [10] J.J. Garcia-Luna-Aceves, "Design and Analysis of Carrier-Sense Multiple Access with Collision Avoidance and Detection," Proc. ACM MSWIM '17, 2017.
- [11] bladeRF 2.0 micro, Nuand LLC, <https://www.nuand.com/bladerf-2-0-micro/> (referred in June 2020)
- [12] USRP B200, Ettus Research, <http://www.ettus.com/all-products/UB200-kit/> (referred in June 2020)



# A Study on System Architecture of Smart Lock Based on Authentication of Door Knocking Motion Using Machine Learning

Kakeru Nakabachi<sup>†</sup>, Yoshitaka Nakamura<sup>‡</sup>, and Hiroshi Inamura<sup>‡</sup>

<sup>†</sup>Graduate School of Systems Information Science, Future University Hakodate, Japan

<sup>‡</sup>School of Systems Information Science, Future University Hakodate, Japan  
{g2119031, y-nakamr, inamura}@fun.ac.jp

**Abstract** - With the development of the Internet of Things (IoT) technologies, smart home systems that provide residents with a more comfortable lifestyle are attracting attention. One of the technologies that make up a smart home system is called Smart lock which allows users to use their smartphones to lock and unlock the doors of their homes instead of a physical key. Smart lock can improve the convenience of locking and unlocking doors. On the other hand, if a resident's smartphone is stolen, there is a risk that a malicious person could break into the user's home., because current Smart lock does not authenticate whether the user who is trying to unlock is the owner of the smartphone. As an approach to this problem, we have been studying personal authentication method using features of the door-knocking motions, which is a natural motion that the user can perform in front of the door. In this paper, we evaluated the performance of the proposed authentication method by calculating F-measure, precision rate, recall rate, and False Acceptance Rate (FAR) for various features that can be obtained from door-knocking motions by using various multi-class classification algorithms in machine learning.

**Keywords:** door knock, smart home, IoT, authentication, smart lock

## 1 INTRODUCTION

In recent years, with the development of Internet of Things (IoT) technologies, smart home systems that provide residents with a more comfortable lifestyle have been attracting a lot of attention. Smart home systems enable residents to operate home appliances such as air conditioners from outside using their own smartphones by installing IoT-compliant home appliances called smart home appliances at home. As one of such smart home services, there is a service called Smart lock that locks and unlocks the user's home door using a smartphone application instead of a physical key. Some Smart lock services have already been commercialized, and some of them can be installed externally on the thumbturn of an existing door to realize a Smart lock service without the need for large-scale construction [1] [2]. In addition to locking and unlocking using a smartphone, which is the main function of the Smart lock service, there are several other functions, such as a hands-free unlocking function that allows users to unlock the door by approaching it using their location information, and a function that allows users to share access privileges with family and friends. These functions improve the convenience of locking and unlocking the door. On the other hand, each

function of the current Smart lock service is performed based on the prior pairing of the Smart lock product and the smartphone, and does not authenticate whether or not the user is the owner of the smartphone. Therefore, if the smartphone is stolen by someone who knows the user's address in advance, or if the smartphone is stolen along with the driver's license with the user's address, etc., there is a risk that a malicious person could break into the user's home. As a countermeasure against these problems, personal authentication near the door is necessary.

Conventional authentication technologies that can be used in front of door include possession-based authentication "what you have", knowledge-based authentication "what you know", and biometric authentication "what you are". Possession authentication is a technology that uses a physical device owned by the user, such as a physical key or an IC card, to authenticate the user. Knowledge authentication is an authentication technique that uses knowledge in user's memory, such as inputting a pre-registered password for authentication. There are two types of biometric authentication that use physical characteristics and behavioral characteristics. Biometric authentication using physical features uses other features of the human body such as fingerprints, face, irises, and so on. Biometric authentication using behavioral features uses features that appear in human behavior such as walking motion, typing motion on a keyboard, or opening motion and closing motion a door. There are advantages and disadvantages to each of these authentication technologies. In the case of possession authentication, users can be authenticated relatively easily using a physical key or IC card. However, if these devices are stolen or lost, personal authentication becomes impossible and there is a risk that someone other than the user will break the authentication. Knowledge authentication is a widely used authentication technology, such as unlocking a smartphone. However, there is a burden on the user's memory, such as the need to remember passwords. In addition, there is a high risk of password theft due to shoulder hacking, etc. Biometric authentication using physical features does not have memory burden such as remembering passwords, since it uses the unique features of the human body. However, because it is nearly impossible to change physical features, it is difficult to deal with the use of replicated physical feature information. In addition, it is necessary to register a part of the body information as a feature using a special device, which is a large psychological burden on the person. Biometric authentication using behavioral features extracts user-specific features from user actions. Therefore, the authentication can-

not be broken unless another person makes the same motion as the user. However, it is difficult to select features to obtain sufficient authentication accuracy. On the other hand, there is a advantage that the features for authentication can be acquired implicitly, such as walking authentication.

In the current commercialized Smart locks, the user approaches the door and locks and unlocks the door at a short distance through Bluetooth, etc., using a smartphone that is paired with the Smart lock device in advance. In other words, the Smart lock device is the keyhole and the smartphone is the key corresponding to the keyhole. Since this is possession-based authentication, it also has the disadvantages of possession-based authentication.

We focus on the door-knocking motion, which is considered to be a behavioral feature that can be performed spontaneously in front of the door and is less burdensome for the motion. In this paper, we propose a door-knock type personal authentication system using biometrics based on personal features detected by door knocking motion. The proposed method is expected to increase the security strength by combining two-factor authentication with biometric authentication using behavioral features, and possession authentication which is performed by existing Smart locks.

The issues of this study can be broadly classified into the following four categories.

- Consideration of system configuration
- Investigation of features that are useful for authentication
- Assessment of resistance to peek-a-boo attacks
- Consideration of extension methods for registered action data

First, it is necessary to study how to handle the acquired motion data and how to configure the system to authenticate the user accurately and correctly. Second, it is necessary to investigate effective features for accurate authentication, which can be obtained from door knocking motions. Third, we need to investigate the resistance of the door knocking motions to being imitated by others when the door knocking is observed by a camera or directly by prying eyes. Fourth, in order to ensure sufficient authentication accuracy with a small amount of registration motion data, it is necessary to consider the extension method of registration motion data.

In addressing these issues, we have investigated the possibility of selecting individual-specific features from door knocking motions in a previous study. In this paper, we report the results of our research on the consideration of the system configuration and the investigation of effective features for authentication.

## 2 RELATED WORK

### 2.1 Authentication Using Behavioral Features

There are some existing authentication methods [3] [4] based on the spatial motion detection using an accelerometer of a

mobile terminal or a wristwatch-type device. Patterned actions such as drawing characters and figures in space, and left hook punching are used as authentication motions using a special device. In addition, researches on rhythm authentication [5] [6] have been proposed. The user registers a patterned rhythm according to the music he or she has listened to beforehand, and the user is authenticated by recalling and reproducing the rhythm pattern at the time of authentication. Thus, there are some biometric authentication methods using behavioral features that require memory burden by asking the user to recall the registration pattern at the time of authentication. For the convenience of users, it is desirable to eliminate these memory burdens. On the other hand, there are methods to authenticate and identify individuals without placing a memory burden on the user by using natural behaviors in daily life, such as walking [7], typing on a keyboard [8], opening and closing doors [9], and rolling up toilet paper [10]. Based on these findings, the proposed method uses the individual features included in the natural knocking motions of daily life, rather than the unnatural knocking rhythm defined for authentication to identify the registration pattern, and removes the memory burden of recalling at the time of authentication.

### 2.2 Authentication of Smart Door Lock Systems

With regard to Smart locks, many studies have been conducted on smart door lock (SDL) systems [11] [12] [13] [14]. Each of these systems proposes its own hardware for door locking and a method for unlocking the door. An SDL system equipped with a touch panel on the door [11] proposes an authentication method in which the user enters a password on the touch panel as authentication when the door is unlocked. In addition, an SDL system [12] has been proposed to unlock door using the one-time password method from a logged-in smartphone application. This method uses knowledge authentication, which results in a large memory burden when unlocking the door. The SDL system [13] that unlock the door automatically when a device with access authorities enters the setting area achieves seamless door unlocking without making the user aware of the operation for authentication. However, when a device with registered access authorities is lost, it is necessary to log in to the web application and delete the registered devices. In addition, a function to authenticate the user and unlock the door using fingerprint authentication [13] has been proposed at the same time. However, since it uses physical features, there is a risk of spoofing if replicated. In addition, there is a significant psychological burden on the user due to the use of specialized devices to obtain fingerprints. An SDL system [14] uses a Passive Infrared Ray (PIR) motion sensor installed on the door to detect the presence of a person in front of the door, and notifies the door administrator to unlock the door. However, the cost of installing the sensor on an existing door and the user who receives the notification has to give permission to unlock the door each time. Sato et al. proposed an SDL system called the Intelligent Doorknob System, which is equipped with palmprint authentication using a Web camera installed on the doorknob [15]. This makes

it possible to unlock the door lock seamlessly without making the user aware of the authentication operation. However, since 150 times SIFT (Scale Invariant Feature Transform) calculations are performed for a single authentication, the current authentication speed is not practical. And the cost of installing a Web camera on the doorknob is also incurred. As described above, many existing SDL systems have problems in terms of user's memory burden, psychological burden, and installation cost. Therefore, an authentication method that requires as little burden on the user as possible and does not require the installation of special equipment is needed.

### 3 PROPOSED SMART LOCK SYSTEM

#### 3.1 Proposed system

In the proposed system, the user is authenticated by using a multi-class classification algorithms in machine learning through a series of processes shown in Fig. 1 after receiving 3-axis acceleration data and 3-axis angular velocity data obtained from the user's door-knocking motion. The Smart lock system installed at home and the smartphone are paired with Bluetooth in advance. The system first investigates whether or not the paired terminals can be authenticated using proximity detection of BLE (Bluetooth Low Energy). By using the distance measurement of the BLE, it is possible to detect the distance between the smartphone and the Smart lock by distinguishing the distance of more than 1 m (Far), about 1 m (Near), and less than 1 m (Immediate). The proposed system permits the authentication by door knock operation only when the BLE detects "Immediate". The door knock motion data acquired by the smartphone sensors at this time is pre-processed with noise removal and feature extraction. Only when the extracted features are determined to be the user's own class by the multi-class classification algorithm, authentication is performed. The Smart lock can only be unlocked when the user is authenticated. Thus, in the proposed system, in addition to the element of possession-based authentication using a smartphone connected to a Smart lock in advance, the authentication can be enhanced by adding an element of biometric authentication using behavioral features by knocking on the door. The proposed system aims to remove the memory burden of remembering the registered pattern by extracting the unique features of the user from the natural door-knocking motion that user performs in daily life.

#### 3.2 Assumed environment

In the proposed system, the feature of door-knocking motion is extracted by the user's smartphone and used for the authentication for unlocking a Smart lock. The proposed system does not install any sensor or device on the door, but uses the sensor built into the smartphone. It is possible to reduce the time and cost to prepare a new dedicated device by using the device that users always carry in their daily lives. The smartphone acquires data from a 3-axis accelerometer and a 3-axis angular velocity sensor. Figure 2 shows the door-knocking motion and the direction of each axis of each sensor.

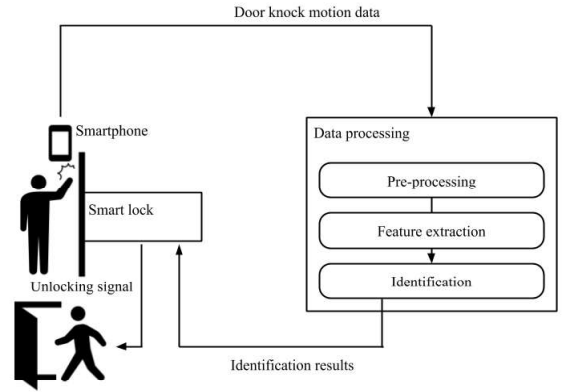


Figure 1: System overview

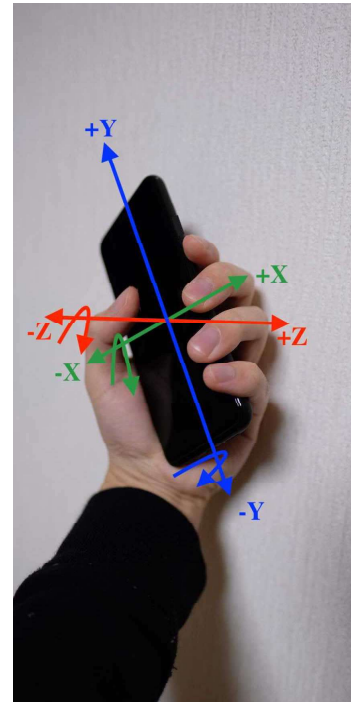


Figure 2: Door knock motion and the axial direction of each sensor

#### 3.3 Pre-processing

In the pre-processing, the 3-axis acceleration data and 3-axis angular velocity data were trimmed to remove the noise except for the actual data. After that, a low-pass filter was applied to each to, to remove the small noise contained in the actual operating data. The result is shown in Fig. 3.

#### 3.4 Feature extraction

From the door-knock motion data obtained in the above environment, the proposed system extracts features that are useful for personal authentication. We have shown that individual-specific features can be extracted from natural door-knocking motions that do not allow for special patterning [16]. The results show that the basic statistics of maximum, minimum, mean, variance, and standard deviation obtained from the 3-axis acceleration and 3-axis angular velocity data and the fea-

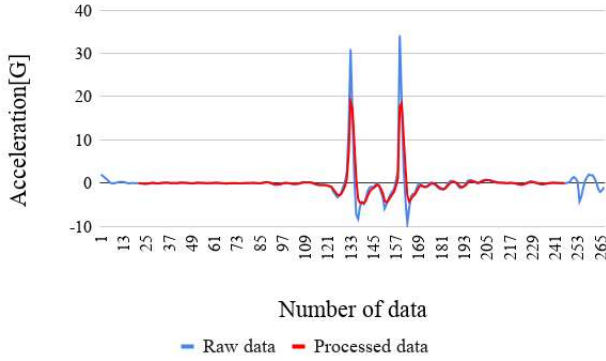


Figure 3: Pre-processing result

tures obtained from the knock peak which is the impact peak at the time of door knock motion appearing in the acceleration data, are effective. However, using the number of knocks and the height and width of each knock as features derived from the knock peaks changes the dimensionality of the feature vectors composed of those features. Therefore, depending on the analysis method of the feature vectors, the same user could be identified as a different user if the number of knocks changed between the registered knocking motion and authentication motions.

In the proposed method, the peak extraction algorithm proposed by Murao et al [17] is used for the knock peak extraction. The results of applying the peak extraction algorithm are shown in Fig. 4. The peak extraction algorithm calculates the average value  $m(t)$  in the past  $\Delta t$  seconds (window) from the current time  $t$  of the acceleration time series data. In contrast, as shown in Fig. 4, a region called the Epsilon tube is established above and below  $m(t)$  with a width of  $m(t) \pm \epsilon$ , and the waveform from once the acceleration value goes out of the Epsilon tube region until it returns to the Epsilon tube region again is extracted as a peak. In the door-knocking motion, we apply this peak extraction algorithm to the z-axis because the door-knocking peak appears directly on the z-axis of acceleration. In this system, the mean, standard deviation and interval of the width and height of each knock peak are extracted as features. In addition, the basic statistics obtained from the 3-axis synthetic acceleration data are also extracted as features, and the frequency spectrum obtained by Fast Fourier Transform (FFT) of the 3-axis synthetic acceleration data is also extracted as features. The types of features that may be used in the proposed system is shown in Table 1. By combining these features, the feature vector patterns shown in Table 2 are finally created and used for identification, respectively. Since each element of the generated feature vectors has a different scale and cannot be treated as equivalent the proposed system normalizes each feature vector so that the mean value of each feature vector is 0 and the variance is 1.

### 3.5 Identification

For identification, we use a multi-class classification algorithm for machine learning. In our previous work [18], we proposed a method of identification using a machine learning anomaly detection algorithm, but it was not sufficiently accu-

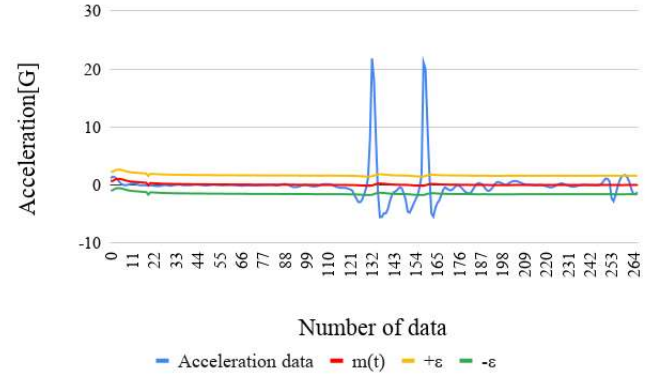


Figure 4: Applying the peak extraction algorithm

Table 1: The types of features

Name	Features	Number of items
3-axis acceleration	3-axis acceleration (max · min · mean · variance · standard deviation)	15
3-axis angular velocity	3-axis angular velocity (max · min · mean · variance · standard deviation)	15
Knock peak	Knock peak width · height · interval (mean · standard deviation)	6
3-axis synthetic acceleration	3-axis synthetic acceleration (max · min · mean · variance · standard deviation)	5
Frequency spectrum	Frequency spectrum of 3-axis synthetic acceleration	100

rate. This method uses an unsupervised learning discriminator for the anomaly detection algorithm. Therefore, it is considered that there was insufficient information to identify the data obtained from the door-knocking motion, which is more difficult to distinguish from other people than the walking motion. By labeling the data to be trained, it is much easier to detect anomalies than unsupervised learning. The classifier learn feature vectors extracted from the motion data by the owner of the Smart lock with the person's label. The feature vectors extracted from the actions by other people are labeled as other people's data and generate a classifier by training it. When a feature vector of a person is given as input to the generated classifier, it is classified as either the person or a stranger the result of the classification is used as the authentication result. The proposed method uses Support Vector Machine (SVM), Logistic Regression, Decision Trees, Random Forest and K Neighbors Classifier as the most commonly used multiclass classification algorithms. In a real system, an appropriate multi-class classification algorithm among these is used. When generating a classifier, if there is little other data for training, the resistance to unknown other data is not enough, and there is a possibility that the user may be mistakenly identified. Therefore, it is assumed that a plurality of other person's motion data sets are collected and used as the other person data for attaching the other person's label at the time of learning.



Table 2: Feature vector patterns

Pattern Number	Features
1	3-axis acceleration · 3-axis angular velocity · Knock peak
2	3-axis acceleration · 3-axis angular velocity
3	3-axis acceleration
4	3-axis angular velocity
5	3-axis synthetic acceleration
6	3-axis synthetic acceleration · 3-axis angular velocity · Knock peak
7	Frequency spectrum
8	3-axis synthetic acceleration · 3-axis angular velocity · Frequency spectrum
9	3-axis acceleration · 3-axis angular velocity · Frequency spectrum
10	3-axis synthetic acceleration · 3-axis angular velocity · Knock peak · Frequency spectrum
11	3-axis acceleration · 3-axis angular velocity · Knock peak · Frequency spectrum

## 4 EXPETIMENTAL EVALUATION

A malicious person with a malicious intent to break into the home may attempt to authenticate with the Smart lock system. Therefore, it is necessary to evaluate the accuracy of classification of the learned data of others. It is also necessary to evaluate the accuracy of the classification of unknown other's data that the classifier has not yet learned to classify. In order to evaluate the effectiveness of the proposed method, the following two evaluations are conducted.

**Evaluation 1** Combination evaluation of best feature vector patterns and multi-class classification algorithms

**Evaluation 2** Evaluation of the effectiveness of features for the feature vector patterns obtained in Evaluation 1

### 4.1 Experiments

In order to evaluate the two evaluations mentioned above, we conducted an experiment with subjects. 7 male university students in their twenties were asked to complete 60 trials each. Subjects were instructed to perform a door-knocking motion with their hands while holding a smartphone, and were instructed to perform natural door-knocking motions in their daily lives, rather than special door-knocking motions. In terms of the number of knocks, we fixed the number of knocks to 2 for all trials, because many subjects knocked twice when they were asked to perform the door-knocking action without instruction from our previous experiment [16].

When knocking on the door, the user was instructed to hold the smartphone in the knocking position before starting the knocking motion, and to hold it still for about 1 second from

Table 3: Main specification of the smartphone

Item	Specification
Device	VAIO Phone JCI VA-10J
OS	Android 5.0.2
Weight	130g
Device size	71.3mm x 141.5mm x 7.95mm
Sensors	3-axis acceleration sensor and 3-axis angular rate sensor
Display size	5-inch
Sampling rate	200Hz

the start of the data measurement until the start of the door-knocking motion and from the end of the door-knocking motion until the end of the data measurement. The above experiment was conducted on the steel door shown in Fig. 5, which is located in the Future University Hakodate and the motion data were obtained. The main specification of the smartphone used in the proposed method are shown in Table 3. An Android application was developed using Kotlin to obtain the door-knocking motion. In order to generate classifiers from feature vectors obtained from motion data, we tuned the parameters of the multi-class classification algorithm to maximize the mean F-measure beforehand, and used 40 out of 60 trial data for training and 20 for testing.



Figure 5: The door used in the experiment

### 4.2 Evaluation method

#### 4.2.1 Evaluation 1

In Evaluation 1, we evaluate whether it is possible to correctly classify the learned person and others, and the accuracy of classification of unknown data that the classifier has not yet learned. We select one of the 7 subject's data as the authenticating user data, and split the data of the remaining 6 subjects into 3 other's data for training and the unknown data. The

Table 4: Results of mean F-measure, mean precision rate and mean recall rate for each feature vector pattern and multi-class classification algorithm

Pattern	SVM			Logistic Regression			Decision Trees			Random Forest			K Neighbors Classifier		
	F-measure	Precision rate	Recall rate	F-measure	Precision rate	Recall rate	F-measure	Precision rate	Recall rate	F-measure	Precision rate	Recall rate	F-measure	Precision rate	Recall rate
1	0.976	0.984	0.969	0.975	0.983	0.967	0.938	0.945	0.933	0.941	0.964	0.923	0.979	0.983	0.976
2	0.968	0.976	0.962	0.966	0.977	0.957	0.903	0.914	0.902	0.933	0.957	0.916	0.976	0.985	0.968
3	0.950	0.962	0.941	0.938	0.958	0.926	0.887	0.889	0.891	0.935	0.956	0.920	0.939	0.954	0.928
4	0.934	0.938	0.933	0.943	0.952	0.937	0.882	0.908	0.868	0.915	0.946	0.895	0.929	0.935	0.925
5	0.941	0.949	0.935	0.933	0.946	0.923	0.881	0.884	0.885	0.886	0.911	0.872	0.871	0.891	0.863
6	0.962	0.974	0.952	0.957	0.971	0.946	0.917	0.942	0.901	0.941	0.973	0.918	0.964	0.973	0.958
7	0.714	0.721	0.715	0.710	0.722	0.709	0.741	0.746	0.740	0.679	0.798	0.672	0.580	0.777	0.609
8	0.963	0.980	0.951	0.970	0.983	0.959	0.893	0.910	0.881	0.894	0.954	0.861	0.961	0.963	0.961
9	0.972	0.984	0.962	0.969	0.983	0.958	0.909	0.925	0.898	0.934	0.966	0.912	0.983	0.984	0.982
10	0.955	0.970	0.944	0.962	0.977	0.949	0.876	0.881	0.876	0.919	0.962	0.893	0.959	0.969	0.952
11	0.978	0.986	0.971	0.977	0.987	0.969	0.940	0.946	0.935	0.953	0.977	0.936	0.983	0.988	0.980

split between other's data and unknown data was done randomly. An evaluation is performed on the processed data by combining 11 feature vector patterns and 5 multi-class classification algorithms. A classifier is generated for each subject by supervised learning in which the data of the user to be authenticated is labeled with authenticating user and the data of another person is labeled with another person. Then, we evaluate whether it is possible to classify the known data correctly using the mean F-measure, the mean precision rate and the mean recall rate of the input data of the known person and others to all the identifiers generated. And to evaluate the classification accuracy of unknown data, we also evaluate the mean False Acceptance Rate (FAR) of untrained unknown data input to the classifier. Finally, to evaluate the best combination of the best feature vector pattern and the multi-class classification algorithm, we use the combination with the maximum value calculated using Eq. 1.

$$meanF-measure + (1 - meanFAR) \quad (1)$$

Precision rate is the percentage of the data that the classifier predicts to be the person's identity, and recall rate is the percentage of the data that the classifier predicts to be the person's identity among the data that the classifier actually is. F-measure is the calculated harmonic mean of these values.

## 4.2.2 Evaluation 2

We already know that the feature vectors are effective for the proposed method among the 11 feature vector patterns, but it is unclear which features are specifically effective for authentication. Therefore, principal component analysis is used to evaluate the importance of each feature and optimize the feature vectors by compressing the number of dimensions of the feature vectors obtained in evaluation 1. In this case, the mean F-measure, the mean precision rate, the mean recall rate and the mean FAR are used to evaluate the feature vector after compression.

## 4.3 Evaluation results and discussion

### 4.3.1 Evaluation 1

The results for each feature vector pattern and the mean F-measure, mean precision rate, and mean recall rate for each

multi-class classification algorithm are shown in Table 4 and the mean FAR results are shown in Table 5. These mean F-measure and mean FAR were evaluated using Eq.1, and it was found that the mean F-measure and the mean FAR were the maximum for the combination of feature vector pattern 4 and Random Forest, which included features extracted from the 3-axis angular velocity data. The mean F-measure was 0.915, the mean precision rate was 0.946, the mean recall rate was 0.895, and the mean FAR was 0.113. The next largest Eq. 1 was the combination of feature vector pattern 2 and Logistic Regression, which included features extracted from the 3-axis acceleration data and 3-axis angular velocity data. The mean F-measure was 0.966, the mean precision rate was 0.977, and the mean recall rate was 0.957, and the mean FAR was 0.196. In both cases, an mean F-measure of more than 0.9 and an mean FAR of less than 0.2 were obtained and the classifier was trained it was found that the classification of the data with high accuracy is possible. Therefore, we can say that it is possible to authenticate individuals using the results of this classification with high accuracy. However, in order for an unknown person to authenticate with the proposed method, the user needs to come close to the Smart lock at less than one meter away from the Smart lock with a smartphone that is paired with the Smart lock installed in the registrant's house. If a malicious person is trying to authenticate the target's home, it is necessary to steal the target's smartphone, know the address of the house, and move to the front door. This takes a lot of work and the possibility of an actual attack is much lower than that of an authentication to unlock the screen of a smartphone. We then need to break through the authentication with the mean FAR probability obtained from the combination of these feature vector patterns and the multiclass classification algorithm. Therefore, the possibility of accidentally authenticating an unlearned stranger is low and the proposed method can be used for authentication because it can classify the trained data with high accuracy. However, it is necessary to confirm whether the same level of accuracy can be obtained when the number of subjects is increased in the future, because these results are based on only 7 subjects.

From the above, the pattern selected as the most suitable among the 11 feature vector patterns were pattern 2, using basic statistics extracted from 3-axis acceleration and 3-axis angular velocity data, and pattern 4 using basic statistics ex-

tracted from 3-axis angular velocity data.

Table 5: FAR results for each feature vector pattern and multi-class classification algorithm

Pattern Number	SVM	Logistic Regression	Decision Trees	Random Forest	K Neighbors Classifier
1	0.325	0.278	0.298	0.350	0.479
2	0.284	0.196	0.244	0.263	0.248
3	0.227	0.198	0.295	0.221	0.236
4	0.302	0.209	0.191	0.113	0.302
5	0.487	0.471	0.490	0.386	0.331
6	0.423	0.398	0.163	0.208	0.445
7	0.328	0.298	0.438	0.131	0.096
8	0.374	0.375	0.274	0.160	0.434
9	0.306	0.209	0.303	0.267	0.468
10	0.389	0.406	0.294	0.217	0.448
11	0.337	0.255	0.256	0.360	0.505

#### 4.3.2 Evaluation 2

For the optimization of the feature vector patterns 2 and 4 obtained in Evaluation 1, the results of mean F-measure, mean precision rate, mean recall rate and mean FAR for each dimensional compression of the vectors by principal component analysis are shown in Table 6. For the calculation of each value, the optimum multi-class classification algorithm for each of the feature vector patterns found by Evaluation 1 was used. The original number of dimensions of this vector is 30, but even if it is compressed to 10 dimensions, each value is almost the same as the value in the original number of dimensions. In addition, even if the number of dimensions after compression is about 6, the accuracy is as good as that of the original number of dimensions, and it is possible that the main features that are useful for authentication are included in the features. Next, we focus on the result of the feature vector pattern 4. The original number of dimensions of this vector is 15 dimensions, but even when compressed to 10 dimensions, we can confirm the same level of accuracy as in the case of the original number of dimensions. It shows that the main features that are valid for authentication are contained in this compressed 10 dimensions.

Table 7 shows the elements of the feature vectors when the number of dimensions of the feature vector patterns 2 and 4 are compressed to less than 10 dimensions, respectively. When we focus on the elements of the compressed feature vector pattern 2, we can see that the features extracted mainly from the acceleration remain. This feature vector pattern 2 was originally a feature vector with the basic statistics extracted from 3-axis acceleration and 3-axis angular velocity. Dimensional compression of feature vectors by principal component analysis shows that acceleration is an effective feature for authentication when features extracted from 3-axis acceleration and 3-axis angular velocity are used in combination. The features extracted from the x-axis and y-axis of the acceleration were found to be effective. In contrast, the feature vector pattern 4 contained the basic statistics extracted from the 3-axis angular velocity as features and it was found that the features were still selected from each axis after compression.

Table 6: The result of compressing feature vectors in each dimension

Pattern 2				
Dimensions	F-measure	Precision rate	Recall rate	FAR
2	0.685	0.675	0.709	0.024
3	0.880	0.912	0.865	0.342
4	0.900	0.921	0.886	0.279
5	0.933	0.946	0.923	0.471
6	0.947	0.965	0.934	0.400
7	0.945	0.962	0.932	0.220
8	0.957	0.971	0.946	0.208
9	0.957	0.971	0.946	0.204
10	0.962	0.973	0.952	0.198

Pattern 4				
Dimensions	F-measure	Precision rate	Recall rate	FAR
2	0.790	0.833	0.771	0.187
3	0.803	0.854	0.785	0.180
4	0.841	0.876	0.822	0.190
5	0.839	0.885	0.818	0.167
6	0.827	0.891	0.801	0.154
7	0.862	0.922	0.832	0.133
8	0.871	0.939	0.841	0.117
9	0.874	0.934	0.845	0.120
10	0.895	0.947	0.865	0.080

Table 7: Elements of feature vectors when the number of dimensions is compressed

Pattern 2	Pattern 4
Accel x-axis mean	Gyro x-axis max
Accel x-axis variance	Gyro x-axis variance
Accel x-axis standard deviation	Gyro x-axis standard deviation
Accel y-axis max	Gyro y-axis mean
Accel y-axis variance	Gyro y-axis variance
Accel y-axis standard deviation	Gyro y-axis standard deviation
	Gyro z-axis max
	Gyro z-axis min
	Gyro z-axis mean
	Gyro z-axis standard deviation

The results show that 3-axis acceleration is a more effective feature for authentication when 3-axis acceleration and 3-axis angular velocity are used together. When only 3-axis angular velocity was used, the accuracy of this method was slightly lower than that of the method using only 3-axis acceleration, but it was found that high accuracy was achieved by using it in combination with 3-axis acceleration. In this evaluation, the relationship between 3-axis acceleration and 3-axis angular velocity was not clear, and it is necessary to investigate it in the future.

## 5 CONCLUSION

In this paper, we proposed a door-knock authentication system using a 3-axis accelerometer and a 3-axis angular velocity sensor as a biometric authentication method using behavioral features that reduces the psychological burden of the user's actions at the time of authentication, without making the user's actions into a pattern. The proposed method is a two-factor authentication method that combines biometric authentication using behavioral characteristics with property authentication using a smartphone paired with a Smart lock.

The results of the experiment with the subjects show that Random Forest is used as a multiclassification algorithm in the prepared combination, and the feature vector pattern 4 using the basic statistics extracted from the 3-axis angular velocity as the feature was found to be the best combination. The mean F-value was 0.915, the mean fit rate was 0.946, the mean reproduction rate was 0.895, and the mean FAR was 0.113. The number of dimensions of the feature vectors in this case was 15, but it was confirmed that the accuracy of the feature vectors was comparable to the original number of dimensions even when compressed to 10.

In the future, we need to evaluate the proposed method through experiments with more subjects and effective feature combinations. In addition, we will use wearable devices such as smartwatches, which have become popular in recent years, to obtain authentication motions, and we consider the selection of appropriate features for acquisition by smartwatches. Also, in the experiments in this paper, the user may remember and reproduce the motions since the knocking motions are obtained continuously. Considering the actual use of the system, there may be a long period of time between the registration of the authentication motion and the authentication, so it is necessary to conduct evaluation experiments, including usability.

## REFERENCES

- [1] Qrio: Qrio Lock Qrio(online), available from <https://qrio.me/smartlock/> (accessed 2019-03-09).
- [2] CANDYHOUSE: Sesami Smart Lock, CANDYHOUSE(online), available from <https://jp.candyhouse.co/> (accessed 2019-03-09).
- [3] Ishihara, S., Ohta, M., Namikata, E. and Mizuno, T.: Individual Authentication for Portable Devices Using Motion of the Devices, *IPSJ Journal*, Vol. 46, No. 12, pp. 2997–3007 (2005). (In Japanese)
- [4] Namikata, E., Ohta, M., Ishihara, S. and Mizuno, T.: An Individual Authentication Method With Arm Movements Using a Wrist Watch Loaded With an Accelerator Sensor, *IPSJ SIG Technical Report*, Vol. 2003, No. 94(2003-HI-105), pp. 21–26 (2003). (In Japanese)
- [5] Ichimura, R., Notomi, K. and Saito, K.: A Rhythm Identification Method for Smart Phones with Peek-a-boo Attack Resistance -Evaluation of authentication accuracy using the main melody of the music-, *DICOMO2013*, Vol. 2013, pp. 230–233 (2013). (In Japanese)
- [6] Kita, Y., Kamizato, K., Park, M. and Okazaki, N.: A study of rhythm authentication using multi-touch operation, *IPSJ SIG Technical Report*, Vol. 2014-UBI-41, No. 19, pp. 1–7 (2014). (In Japanese)
- [7] Konno, S., Nakamura, Y., Shiraishi, Y. and Takahashi, O.: Improvement of accuracy based on multi-sample and multi-sensor in the gait-based authentication using trouser front-pocket sensors, *International Journal of Informatics Society (IJIS)*, Vol.8, No.1, pp.3-13, (2016).
- [8] Ito, S., Shiraishi, Y. and Konno, S.: A Method for Personal Authentication by Using Wrist-mounted Sensors Based on Characteristics of Keystroke Action, *DICOMO2016*, Vol. 2016, pp. 1165–1171 (2016). (In Japanese)
- [9] Mitsukude, Y., Hayashi, K., Ishida, S., Tagashira, S. and Fukuda, A.: Proposal and Initial Evaluation of Human Identification Based on Door Opening/Closing Operations, *IPSJ SIG Technical Report*, Vol. 2019-UBI-61, No. 32, pp. 1–6 (2019). (In Japanese)
- [10] Kurahashi, M., Murao, K., Terada, T. and Tsukamoto, M.: A System for Identifying Toilet User by Characteristics of Paper Roll Rotation, *Proc. of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQutous 2016)*, pp. 282-283 (2016).
- [11] Park, Y. T., Sthapit, P. and Pyun, J.: Smart digital door lock for the home automation, *Proc. 2009 IEEE Region 10 Conference(TENCON2009)*, pp. 1–6 (2009).
- [12] Dhondge, K., Ayinala, K., Choi, B. and Song, S.: Infrared Optical Wireless Communication for Smart Door Locks Using Smartphones, *Proc. 2016 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pp. 251–257 (2016).
- [13] Hadis, M. S., Palantei, E., Ilham, A. A. and Hendra, A.: Design of smart lock system for doors with special features using bluetooth technology, *Proc. 2018 International Conference on Information and Communications Technology (ICOIACT2018)*, pp. 396–400 (2018).
- [14] Aman, F. and Anitha, C.: Motion sensing and image capturing based smart door system on android platform, *Proc. 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 2346–2350 (2017).
- [15] Sato, K., Noma, Y., Kashima, M. and Watanabe, M.: A Study on the Development of an Intelligent Door Knob System with Palmprint Authentication, *MIRU2011*, Vol. 2011, pp. 580–585 (2011). (In Japanese)
- [16] Nakabachi, K., Nakamura, Y. and Inamura, H.: Door

- knock type person authentication method for smart lock system, *DICOMO2019*, Vol. 2019, pp. 1433–1440 (2019). (*In Japanese*)
- [17] Murao, K., VanLaerhoven, K., Terada, T. and Nishio S.: A Method for Context Awareness using Peak Values of Sensors, *IPSJ SIG Technical Report*, Vol. 2009-UBI-22, No. 11, pp. 1–8 (2009). (*In Japanese*)
- [18] Nakabachi, K., Nakamura, Y. and Inamura, H.: Improvement of door knock type authentication method for smart door lock system, *IPSJ SIG Technical Report*, Vol. 2020-CSEC-88, No.45, pp.1-7 (2020). (*In Japanese*)





# A Proposal of Personal Authentication Method Based on Eye Movement Trajectory with Fixation and Saccade Features

Takumi Fujimoto\* and Yoh Shiraishi\*\*

\*Graduate School of Systems Information Science, Future University Hakodate, Japan

\*\*School of Systems Information Science, Future University Hakodate, Japan  
{g2119039, siraisi}@fun.ac.jp

*Abstract* – Password authentication and biometric authentication have become popular as personal authentication technology for mobile terminals. However, these technologies have weaknesses regarding security. In password authentication, authentication information can be leaked when one person surreptitiously looks at another's password. In biometric authentication, it is difficult to deal with impersonation when authentication information has been forged. We focus on user's eye movement to overcome these weaknesses. There are researches which use unconscious eye movement on visual stimuli for authentication [1], [2]. In these methods, it is difficult to update the registered authentication information. There are researches which use conscious eye movement for inputting password [3], [4]. These methods have the risk that its authentication information can be guessed. Our study proposes a personal authentication method based on eye movement trajectory when users draw with their eyes on the display of a mobile terminal. The proposed system performs authentication based on the shape of the user's eye movement trajectory and authentication based on its drawing features. We used fixation and saccade features for classifying users, and evaluated whether these features are effective for the proposed method. The experimental result showed the proposed method improved classification accuracy when compared to our previous method. Next, we used One Class SVM and Isolation Forest for personal identification to examine a learning algorithm. The experimental result showed that Isolation Forest was effective for personal identification. In future work, we will consider a method to solve the lack of learning data. In addition, we need to investigate other local eye movement features and learning algorithms.

*Keywords:* Personal authentication; Eye movement trajectory; Drawing feature; Classification of users; Error detection;

## 1 INTRODUCTION

In recent years, many people perform personal authentication on web services and applications on mobile terminals. If the authentication information is leaked, there is a risk of it being used fraudulently. It is important to improve the security level of authentication on mobile terminals.

Password and biometric authentication are popular as the authentication on mobile terminals. Password authentication is authentication that uses the user's knowledge such as passwords and PINs. Password authentication does not require a physical object such as an IC card or key as authentication information.

However, authentication information can be leaked when someone looks over another person's shoulder in a public space. In addition, there is a risk that authentication information can be guessed.

Biometric authentication is authentication that uses a part of the body (physical features) and human behavior (behavioral features) as authentication information. Biometric authentication is robust against over-the-shoulder information theft and incurs less burden. However, in authentication based on physical features, there is a risk that the physical features registered as authentication information may be forged. It is difficult to deal with impersonation if authentication information is forged, because information cannot be consciously altered. In authentication based on behavioral features, it is difficult to forge the authentication information. However, this authentication information cannot update.

The weakness of password authentication is the ease of leakage of authentication information. In addition, this authentication information be guessed easily. The weakness of biometric authentication is the difficulty of dealing with impersonation. These weaknesses must be overcome in order to perform secure authentication on mobile terminals. There are many studies using physical or behavior features for authentication [3-8]. In addition, as examples of methods that are robust against over-the-shoulder information theft, there are studies using user eye movement for authentication [1], [3], [9-11]. We think user's eye movement prevents leakage of authentication information because it is difficult for others to observe eye movement. In addition, we think that it is difficult to guess the eye movement information. Authentication information is updatable by the user eye movement because user's eye movement can be consciously reproduced.

This study proposes a personal authentication method using eye movement trajectory, that is the trajectory drawn by the user's eye movement on a mobile terminal (eye movement trajectory). The goal of this study is to realize personal authentication that solves the problems with password and biometric authentication. In our previous research, we investigated the features for trajectory classification based on the shape of the eye movement trajectory, and feature values of drawing features [12]. As feature values for the shape of trajectory, we investigated features that can be extracted from images and coordinates data of trajectory. The results show that coordinates data are effective for the proposed method. We extracted global drawing features from eye movement trajectory for classifying users, but sufficient accuracy was not obtained. In this paper, we reexamined

drawing features to improve the accuracy of personal authentication. In addition, we examined the learning algorithm for building an authentication model.

In section 2, we explain related works using behavioral features or eye movement for authentication. In section 3, we explain the details of the proposed method for personal authentication. In section 4, we explain the experiment to investigate the local eye movement features used in the proposed method, and show the effectiveness of the error detection algorithm for the proposed method. In section 5, we conclude this paper.

## 2 RELATED WORK

### 2.1 Authentication based on behavioral features

There are studies using keystroke for authentication [5], [6]. Nakakuni et al propose a method that uses features of keystroke dynamics when users enter their surname for authentication [5]. Zhou et al. used keystroke acoustic and keystroke dynamics features for authentication [6]. In this method, they calculate MFCC (Mel-Frequency Cepstral Coefficients) by the key acoustics.

There are studies using walking features for authentication [7], [8]. Li et al. use walking features while holding a mobile phone for authentication [7]. Musale et al. extract leg and arm movement features with smartwatch and smartphone during walking and use these features for authentication [8].

There are studies using the features of smartphone operation for authentication [9], [10]. Salem et al. use keystrokes as a second authentication factor when performing authentication with a touchscreen terminal [9]. In this method, they use features such as the timing and position of presses on the keyboard for authentication. Ito et al. have proposed an authentication method based on the features of the flick input method on smartphones [10]. Features such as flicking and shaking of the terminal during text input are used for continuous authentication.

These methods are difficult to update authentication information because these use unconscious habits and patterns.

### 2.2 Authentication based on eye movement features

There are studies using unconscious or conscious eye movement for authentication [1], [3], [9-11].

Kinnunen et al. propose an authentication method using features of unconscious eye movement when users are viewing a video [1]. Ma et al. propose an authentication method using eye movement and head movement on random visual stimuli [2].

In these methods, it will be difficult to update the registered authentication information.

As studies using conscious eye movement for authentication, there are studies performing authentication using password by eye movement [3], [4]. In addition, there is a study that performs authentication by having the user draw with their eye movement trajectory [11]. De Luca et al. propose a password authentication method by gazing at PIN keys

shown on a display [3]. Khamis et al. propose a personal authentication method that uses multimodal passwords with touch input and gaze direction (e.g., left-3-right-4) [4]. In contrast, Mukai et al. have users draw a specified character with their eyes, and use the features of the eye movement trajectory for classifying users [11]. This study uses a character or symbol that can be registered for authentication information.

We think that in the methods proposed in [3,4], authentication information can be easily guessed by attackers. In the method used in [11], the characters that can be used as authentication information are limited because there are a few characters that can be used as authentication information. Therefore, the method has the risk of authentication information being guessed.

## 3 METHOD

### 3.1 Goal of our study

The goal of this study is to propose a personal authentication method based on the user's eye movement trajectory to overcome the weaknesses of password and biometric authentication.

The proposed method consists of authentication based on the shape of the user's eye movement trajectory and authentication based on personal features when the user draws the eye movement trajectory (drawing features). The shape of the eye movement trajectory is defined by users and can be consciously updated by the users. We think that defining the shape of trajectory by users themselves makes difficult to guess the authentication information. We introduce the authentication based on drawing features to makes the proposed method more robust against impersonation.

The goal of this paper is to extract local eye movement features and investigate learning algorithms suitable for 1:1 authentication.

### 3.2 Proposed system

Figure 1 shows an overview of the proposed system

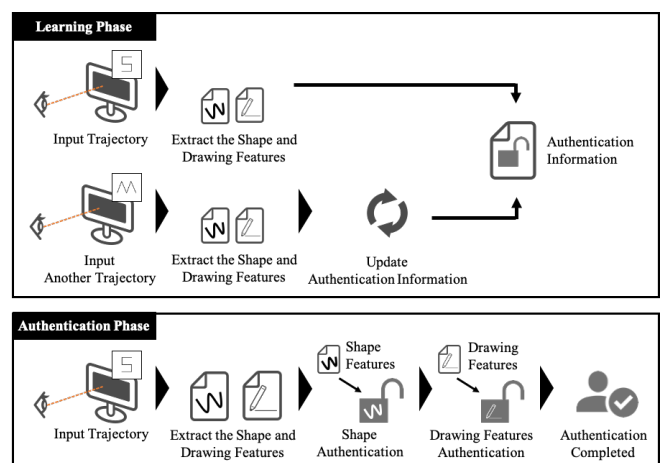


Figure 1 An overview of the proposed system

The proposed system consists of a learning phase and an

authentication phase. The learning phase is the phase of creating the authentication information with multiple entries. Users input and register multiple eye movement trajectories. The shape and drawing features of the eye movement trajectory are extracted and registered as authentication information. In addition, users can update the authentication information by entering the new eye movement trajectory multiple times. The authentication phase is the phase which authentication is performed. In this phase, users enter the eye movement trajectory registered on the learning phase for authentication. First, the authentication based on the shape is performed. Next, the authentication based on drawing features is performed. The authentication of the proposed method is completed by success of these two steps authentication.

### 3.3 Research tasks and approaches

The main research tasks of this paper are as follows:

- **Research task 1:** Selecting a measurement device to use in the proposed method.
- **Research task 2:** Investigating features which are effective for authentication based on shape.
- **Research task 3:** Investigating features which are effective for authentication based on drawing features.
- **Research task 4:** Investigating learning methods for 1:1 authentication.

The approaches to these research tasks are as follows:

- **Approach to research task 1:** We use a contactless type device. As measurement devices, there are contact type devices and contactless devices [13]. We think contact type devices impose a burden, such as a sense of unfamiliarity and blocking of sight for users who do not normally wear glasses. If users use contactless type devices, users need not to wear the devices.
- **Approach to research task 2:** We use features that can be extracted from the coordinates data of the eye movement trajectory. In our previous study [12], we investigated features that are effective for estimating the shape of trajectory by using coordinates data and images. The F-measure of using coordinates data was 0.96 and the F-measure of using images was 0.72. The experimental results showed that coordinates data was effective.
- **Approach to research task 3:** We use fixation and saccade features for classifying users. In our previous study, we classified users by using the amount of change in the coordinates of all frames of the eye movement trajectory during drawing. However, some of the features were not effective for classifying users. In this paper, we extract fixation and saccade features from the user's eye movement trajectory.
- **Approach to research task 4:** We use an error detection algorithm. Error detection is a method that learns only normal data and identifies whether unknown data is normal or error. The proposed method performs 1:1 authentication. This authentication uses only user's data for learning and identifies the user. We assume that the proposed method is applied to

mobile terminals retained by people. Therefore, we use the error detection algorithm for the proposed method.

In this paper, we focus on research tasks 3 and 4. We evaluate effectiveness of fixation and saccade feature for classifying users. In addition, we use One Class SVM (Support Vector Machine) and Isolation Forest for personal identification to evaluate the effectiveness of error detection algorithms for personal identification.

### 3.4 Measurement device and data

In this study, we use a contactless type device as a measurement device. We use the Tobii Pro Tx-300 (Figure 2) for eye tracking. The positions of the line of sight on the screen are recorded at about 60 Hz. Subjects sit 60 cm away from the screen. We instructed subjects to draw eye movement trajectory within the range of the screen and not to move their heads during drawing eye movement trajectory. Only the shape of the trajectory to be drawn was indicated, and the size was not specified. During the measurement, the trajectory drawn by the subject's line of sight is not displayed on the screen, nor is there a guide for drawing it. The measurement environment is shown in Figure 3. An example of the collected data is shown in Figure 4.



Figure 2 Measurement equipment (Tobii Pro Tx-300)



Figure 3 Measurement environment

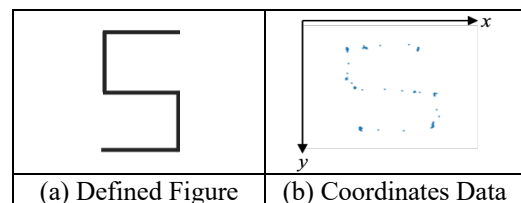


Figure 4 An Example of collected data

The coordinates data consists of coordinates and their corresponding time stamps chronologically recorded. In the proposed method, this data is preprocessed. We extract features used for authentication based on the shape or drawing features from the preprocessed data.

### 3.5 Preprocessing data

We think shakes of eye movement and fixations that occur in the raw coordinates data pose a hindrance to the authentication based on shape. On the other hand, we think shakes of eye movement and fixations are effective for the authentication based on drawing features. Therefore, we perform preprocessing before each authentication.

First, we divide all the frames of the drawn trajectory in chronological order. Next, we calculate the average of all coordinates in each partitioned area and generate the average coordinate data. A smaller number of divisions can remove shakes of eye movement and maintain the approximate shape information of trajectory.

### 3.6 Investigation of the features for the proposed method

#### 3.6.1. Investigation of drawing features for personal identification

In this paper, we use fixation and saccade features for personal identification. Fixation is a local eye movement that is performed to fix the direction of sight. Saccade is a local eye movement that is directed from one point to another [14]. We think personal features appear in these eye movements. The proposed method extracts these eye movement features.




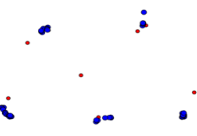

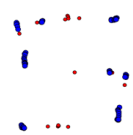
#### 3.6.2. Extracting fixation and saccade features

First, we explain about extracting fixation features. We detect fixation points from average coordinates by window sliding. The points where the coordinates are crowded in each window sliding is fixation points. In this paper, points where five or more coordinates are crowded were detected as fixation points. We extract fixation features for each fixation point. The fixation occurs multiple times during drawing trajectory by the eye movement. Therefore, we calculate the features from all fixations detected.

Second, we extract saccade features. On the eye movement trajectory, fixation is expected to occur at the turning points. In addition, saccades are generated when moving the line of sight between the turning points. It is thought that the user's drawing trajectory consists of the repetition of fixation and saccade. In order to extract the saccade, we focus on the fixation. We calculate average coordinates for each fixation point and calculate coordinates data consisting of only fixation points. Next, we extract saccade features by calculating the amount of change between two consecutive frames of the average coordinates data for each fixation point. The saccades are occurred multiple times during the drawing the eye movement trajectory, as well as the fixation. Therefore, we calculate the features used in the proposed method from all saccades detected during drawing an eye movement trajectory.

We analyzed whether the fixation points could be detected by using the proposed method. We extracted fixation points from the average coordinates data of four trajectories. Table 1 shows the results of the analysis.

Table 1 Fixations extracted from trajectory

Trajectory	Fixations
	
	
	

Blue points are where fixations were observed. Blue points are crowded at the starting and, finishing point of the trajectory and turning points.

#### 3.6.3. Features used for classifying users

Table 2 shows features used for the proposed method and our previous study [12]. The proposed method uses fixation and saccade features are used as local features. Drawing time is the time from the start to the end of the drawing. Fixation time includes the maximum and average times that occurred fixation at each point. Variance of fixation includes the maximum, minimum and average variance that occurred fixation at each point. Standard deviation of fixation is calculated in the same way as variance of fixation. The features for fixation and saccade shown in Table 2 are regarded as local features.

Table 2 Features used for experiments

Features	Previous study [12]	Proposed method
Amount of change in average coordinates data	○	×
Variance of x and y coordinates	○	○
Standard deviation of x and y coordinates	○	○
Drawing time	○	○
Fixation time (Maximum, Average)	×	○
Variance of fixation (Maximum, Minimum, Average)	×	○
Standard deviation of fixation (Maximum, Minimum, Average)	×	○
Number of occurrences of fixation	×	○
Speed of saccade in the x and y directions (Maximum, Minimum, Average)	×	○
Number of occurrences of saccade	×	○

### 3.7 Investigation of learning algorithm

There are studies using error detection algorithms for personal authentication [10], [15]. These studies use One Class SVM or Isolation Forest. One Class SVM is an error detection algorithm that learns only normal data in SVM and identifies whether unknown input data is normal or error data. Isolation Forest is an error detection algorithm that detects error data by repeatedly selecting features and dividing points of data classes. In this paper, we use these algorithms for personal identification.

## 4 EVALUATION

### 4.1 Classifying users using local features

We performed classification of five subjects by using Random Forest to evaluate the effectiveness of fixation and saccade features for classifying users. We instructed the subjects to draw the trajectory shown in Figure 5 30 times. Table 2 shows the features used for classifying users in the proposed method and our previous study [12]. We classify users and calculate the F-measure to evaluate classification accuracy by 10-fold cross-validation. We calculate variable importance to investigate the effective features for classifying users. Table 3 shows the environment of the experiments. Figure 6 shows the result of classification.

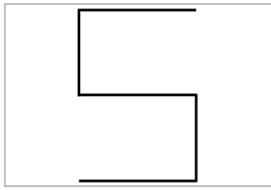


Figure 5 Trajectory used for the experiments

Table 3 Environment of the experiments

Items	Specification
CPU	Intel Core i5 2.4GHz
OS	macOS Catalina10.15.2
Language	Python3.4.5
Library	scikit-learn0.18.1

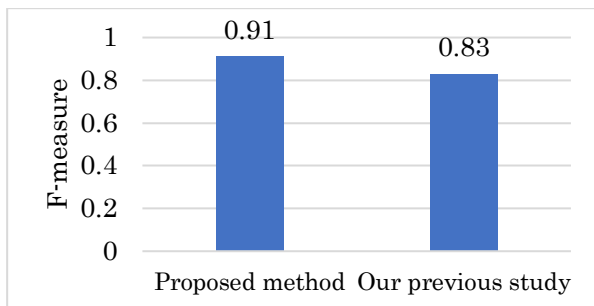


Figure 6 Result of classifying users

The experimental result shows that the F-measure of the proposed method was 0.91 and the F-measure of our previous study was 0.83.

Table 4 shows the top 10 features of the proposed method with variable importance. The variable importance with re-

lated on saccade is high. These results show that saccade features contribute to the classifying users. However, the variable importance of fixation is low and it is not included in the top 10 features shown in Table 4. In this paper, fixation points are defined as the points where five or more coordinates are crowded. Fixations are not detected if the time of fixation is short. Therefore, we think that the variable importance of fixation was low because we could not extract the personal features of fixation.

Table 4 Variable importance of features used to classify users

Features	Variable importance	
	Proposed method	Our previous study
Standard deviation of x coordinates	1.99	0.56
Variance of x coordinates	1.69	0.30
Minimum speed of saccade in x direction	1.48	-
Average speed of saccade in x direction	0.62	-
Drawing time	0.61	0.15
Maximum speed of saccade in x direction	0.55	-
Average speed of saccade in y direction	0.45	-
Minimum speed of saccade in y direction	0.44	-
Variance of y coordinates	0.36	0.17
Maximum speed of saccade in y direction	0.35	-

These results showed that the features of local eye movement (Minimum speed of saccade in x direction, Average speed of saccade in x direction, Maximum speed of saccade in x direction, Average speed of saccade in y direction, Minimum speed of saccade in y direction, Maximum speed of saccade in y direction) are effective for classifying users.

### 4.2 Personal identification using error detection algorithms

We use One Class SVM and Isolation Forest for personal identification to evaluate the effectiveness of the error detection algorithms for the proposed method. We instructed five test subjects to draw the trajectory shown in Figure 5 30 times. We use the features shown in Table 2 for this experiment. We evaluate the identification accuracy by F-measure, FAR (False Acceptance Rate), and FRR (False Rejection Rate). FAR is the probability of mistakenly identifying others (non-users) as authentication users. FRR is the probability of mistakenly identifying users as others. The result of identification is shown in Table 5.

Table 5 Result of identification (F-measure, FAR and FRR)

Algorithms	F-measure	FAR	FRR
One Class SVM	0.23	0.00	0.87
Isolation Forest	0.73	0.08	0.27



In the case of using One Class SVM, the F-measure was 0.23, FAR was 0.00, and FRR was 0.87. In the case of using Isolation Forest, the F-measure was 0.73, FAR was 0.08, and FRR was 0.27. These results showed that the learning models could reject others and accept the user with a high degree of probability. This result showed that users were accepted with relatively high accuracy in the case of Isolation Forest. We think that we were unable to build a learning model which is effective for personal identification, because there was not sufficient learning data. The results of the evaluation suggested that Isolation Forest was effective as an algorithm for personal identification in our study.

## 5 CONCLUSION

This study proposes a personal authentication method using the user's eye movement trajectory to solve the weaknesses of password and biometric authentication. This method performs authentication based on the shape of the user's eye movement trajectory, and authentication based on drawing features. In this paper, we extract fixation and saccade features and classify users to investigate features that are effective for classifying users. In addition, we conducted an experiment that identified users by using representative error detection algorithms in order to investigate the learning algorithm for the proposed method.

First, we performed classifying users by using fixation and saccade features. The experimental result showed that the F-measure of the proposed method was 0.91 and the F-measure of the previous study was 0.83. Therefore, it was suggested that fixation and saccade were effective for classifying users. The saccade variable importance was high. The results showed that local eye movement features were effective for classifying users.

Next, we identified users by One Class SVM and Isolation Forest to investigate a learning algorithm. We evaluated the identification accuracy by F-measure, FAR, and FRR. The experimental result showed that the accuracy was high when using Isolation Forest for personal identification. Therefore, it was suggested that Isolation Forest was effective as the algorithm for the proposed method.

As the future tasks, we need to solve the lack of learning data. In addition, we need to consider other local eye movement features and learning algorithms.

## ACKNOWLEDGEMENTS

Part of this work was carried out under the Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University.

## REFERENCES

[1] T. Kinnunen, F. Sedlak and R. Bednarik, "Towards Task-Independent Person Authentication Using Eye Movement Signals," Proceedings of the 2010 ACM Symposium on Eye-Tracking Research & Applications, ETRA'10, pp.187-190 (2010).  
 [2] Z. Ma, X. Wang, R. Ma, Z. Wang and J. Ma, "Integrating Gaze Tracking and Head-Motion Prediction for Mobile Device Authentication: A Proof of Concept," Sensors, Vol.18,

No.9 (2018).

[3] A. De Luca, R. Weiss and H. Drewes, "Evaluation of Eye-Gaze Interaction Methods for Security Enhanced PIN-Entry," Proceedings of the 19<sup>th</sup> Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, OZCHI'7, pp.199-202 (2007).  
 [4] M. Khamis, F. Alt, M. Hassib, E. Zezschwitz, R. Hasholzner and A. Bulling, "GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices," Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp.2156-2164 (2016).  
 [5] M. Nakakuni and H. Dozono, "User Authentication Method for Computer-based Online Testing by Analysis of Keystroke Timing at the Input of a Family Name," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), pp.71-76 (2018).  
 [6] Q. Zhou, Y. Yang, F. Hong, Y. Feng and Z. Guo, "User Identification and Authentication Using Keystroke Dynamics with Acoustic Signal," 2016 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), pp.445-449 (2016).  
 [7] H. Li, J. Yu and Q. Cao, "Intelligent Walk Authentication: Implicit Authentication When You Walk with Smartphone," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp.1113-1116 (2018).  
 [8] P. Musale, D. Baek, N. Werellagama, S.Woo and B. Choi, "You Walk, We Authenticate: Lightweight Seamless Authentication Based on Gait in Wearable IoT Systems," IEEE Access, Vol.7, pp.37883-37895 (2019).  
 [9] A. Salem, D. Zaidan, A. Swidan and R. Saifan, "Analysis of Strong Password Using Keystroke Dynamics Authentication in Touch Screen Devices," 2016 Cybersecurity and Cyberforensics Conference (CCC), pp.15-21 (2016).  
 [10] S. Ito and Y. Shiraishi, "A Proposal of A Method for Continuous Authentication Focusing on Characteristics of Flick Input System On Smartphones", Proceedings of 25th Multimedia Communication and Distributed Processing Workshop, Vol.2017, pp.1-8 (2017) (*in Japanese*).  
 [11] H. Mukai and T. Ogawa, "Feature Extraction of Eye-gaze Path for Personal Authentication," IPSJ Journal Digital Contents (DCON), Vol.4, No.2, pp27-32 (2018) (*in Japanese*).  
 [12] T. Fujimoto and Y. Shiraishi, "An Examination of Personal Authentication Method Using Shape of Gaze Trajectory and Drawing Features," Proc. Multimedia, Distributed, Collaborated and Mobile Symposium of IPSJ, pp.1423-1432 (2019) (*in Japanese*).  
 [13] Tobii Pro Mechanism of Eye Tracker, tobii pro, <https://www.tobiiipro.com/ja/service-support/learning-center/eye-tracking-essentials/how-do-tobii-eye-trackers-work/> (accessed 2019-05-06).  
 [14] K. Ukai, Eye Movement: Characteristics and Method of Measurement, Japanese Journal of Optics, Vol.23, No.1, pp.2-8 (1994) (*in Japanese*).  
 [15] K. Watanabe, M. Nagatomo, K. Aburada, N. Okazaki and M. Park, "Gait- Based Authentication using Anomaly Detection with Acceleration of Two Devices in Smart Lock," Proceedings of the 14th International Conference on Broad-Band Wireless Computing, Communication and Applications (BWCCA- 2019), Lecture Notes in Networks and Systems 97, Springer, pp.352-362 (2019).



# Hybrid CSMA/CA for Sub-1 GHz Frequency Band Coexistence of IEEE 802.11ah and IEEE 802.15.4g

Yukimasa Nagai<sup>†\*</sup>, Jianlin Guo<sup>‡</sup>, Takenori Sumi<sup>†</sup>, Philip Orlik<sup>‡</sup>, and Hiroshi Mineno<sup>\*</sup>

<sup>†</sup>Mitsubishi Electric IT R&D Center, Japan

<sup>‡</sup>Mitsubishi Electric Research Laboratories, USA

<sup>\*</sup> Graduate School of Science and Technology, Shizuoka University, Japan  
Nagai.Yukimasa@ds.MitsubishiElectric.co.jp

**Abstract** - As more and more wireless technologies have been developed to support emerging IoT applications, the coexistence of heterogeneous wireless technologies presents challenges. IEEE 802.15.4g and IEEE 802.11ah are two of such wireless technologies specified for outdoor IoT applications. Due to the constrained spectrum allocation in the Sub-1 GHz (S1G) band, these two types of devices may be forced to coexist, i.e., share frequency spectrum. To investigate coexistence behavior of 802.15.4g and 802.11ah, we first identify coexistence issues using our newly developed NS-3 based S1G band coexistence simulator. Accordingly, we propose a hybrid CSMA/CA mechanism for 802.15.4g to address the identified coexistence issues. The conducted performance analysis shows that the proposed hybrid CSMA/CA improves 802.15.4g performance without degrading 802.11ah performance. The hybrid CSMA/CA also maintain overall 802.11ah packet latency.

**Keywords:** Wireless coexistence, hybrid CSMA/CA, Sub-1 GHz band, WLAN, WPAN.

## 1 Introduction

The Internet of Things (IoT) applications are rapidly growing. A broad range of wireless technologies have been developed to cater the diverse applications. As heterogeneous wireless technologies are emerging, coexistence becomes a critical issue to be addressed. IEEE 802.11ah, marketed as Wi-Fi HaLow, is the first 802.11 standard designed to operate in the Sub-1 GHz (S1G) band. IEEE 802.15.4g, marketed as Wi-SUN, also operates in the S1G band for outdoor IoT applications. The unlicensed spectrum allocation is limited, especially in the S1G band compared with other 2.4 GHz band. For example, Japan only allocates 7.6 MHz spectrum in 920 MHz band for active radio devices in the ARIB STD-T108 (20 mW) [1]. The constraint spectrum allocation indicates that 802.11ah devices and 802.15.4g devices may be forced to coexist, i.e., share frequency spectrum. In addition, 802.11ah network and 802.15.4g network can have thousands of nodes. Both technologies have communication range of 1000 meters for IoT applications. These features significantly increases the coexistence potential. Therefore, ensuring harmonious coexistence of these two wireless technologies is important.

802.11ah mandates the support of 1 MHz channel, which is much narrower than the 20 MHz channel for conventional 802.11 in the 2.4 GHz band. As a result, the existing

coexistence technologies designed for the 2.4 GHz band may not be suitable for the coexistence of 802.11ah and 802.15.4g in the S1G band. Therefore, the coexistence of 802.11ah and 802.15.4g needs to be further investigated. Accordingly, IEEE New Standards Committee and Standard Board formed IEEE 802.19.3 Task Group in December 2018 to develop an IEEE 802 standard for the coexistence of 802.11ah and 802.15.4g in the S1G frequency band [2]. Authors of this paper have been leading this standard development.

[3] proposes a prediction based self-transmission control method for 802.11ah to ease its interference impact on 802.15.4g. [4] introduces  $\alpha$ -Fairness ED-CCA method for 802.11ah to mitigate its interference on 802.15.4g caused by its higher ED threshold. To address the interference caused by the faster backoff of 802.11ah, [4] also proposes Q-Learning based backoff mechanism for 802.11ah to avoid interfering with 802.15.4g packet transmission process. However, these coexistence technologies improve the performance of 802.15.4g at the expense of 802.11ah. This paper aims to develop coexistence technologies that improve 802.15.4g performance without degrading 802.11ah performance. We first evaluate coexistence behavior and identify coexistence issues by using the developed S1G band coexistence simulator. We then propose a hybrid CSMA/CA mechanism for 802.15.4g to achieve better coexistence with 802.11ah.

The rest of this paper is organized as follows. Section 2 presents related work. Section 3 evaluates coexistence behavior and issue of 802.11ah and 802.15.4g. We introduce the proposed hybrid CSMA/CA mechanism in Section 4. In Section 5, we introduce our S1G band coexistence simulator. Performance evaluation of the hybrid CSMA/CA mechanism is conducted in Section 6. Then, we conclude our work.

## 2 Related Work

There are existing coexistence technologies developed for conventional 802.15.4 to address its coexistence with 802.11 in the 2.4 GHz band. [5] proposes a decentralized approach to mitigate interference by adaptively adjusting energy detection (ED) threshold. [6] proposes an adaptive backoff mechanism to survive coexistence with 802.11. [7] designs a cooperative busy tone method via a special device to enable 802.11 to be aware of 802.15.4 transmission. [8] proposes a hybrid device to coordinate 802.11 and 802.15.4 transmissions. [9] proposes an adaptive interference mitigation scheme for 802.15.4 to control its frame length based on the measured 802.11 interference via a hybrid device.

Table 1: The majority of available performance evaluation, and conventional coexistence researches.

Reference	Year	Target System	Band	Objective	Validation Tool
<i>This article</i>	2020	11ah & 15.4g	Sub-1 GHz	delivery rate and latency	ns-3
J. Guo, P. Orlik [3]	2017	11ah & 15.4g	Sub-1 GHz	delivery rate and latency	ns-3
Y. Liu, J. Guo et al.[4]	2018	11ah & 15.4g	Sub-1 GHz	delivery rate and latency	ns-3
W. Yuan et al. [5]	2010	11b & 15.4	2.4 GHz	throughput	OPNET
E.D.N Ndihi et al. [6]	2016	11 & 15.4	2.4 GHz	delivery rate	MATLAB
X. Zhang, et al. [7]	2011	11 & 15.4	2.4 GHz	analytical model, throughput	analytical, ns-2
J.Hou et al. [8]	2009	11 & 15.4	2.4 GHz	delivery rate	experiments
J.W. Chong et al. [9]	2015	11 & 15.4	2.4 GHz	throughput	analytical
B. Badihi et al. [10]	2013	11ah & 15.4	Sub-1 GHz	throughput and energy consumption	OMNeT++
R. Ma et al. [11]	2017	11b & 15.4	2.4 GHz	analytical model, throughput	analytical & unknown simulator

Before the work in [3] and [4], to the best of our knowledge, no other existing work addresses the coexistence of 802.11ah and 802.15.4g in the S1G band. The related studies are done either for 802.11ah or 802.15.4g only. [10] compares performance of 802.11ah and conventional 802.15.4 in the S1G band. The results reveal that 802.11ah network achieves higher channel efficiency than 802.15.4 network. [11] investigates the coexistence issues of 802.11b and 802.15.4 in the 2.4 GHz band. It shows that 802.11b can significantly interfere with 802.15.4. However, our investigation shows that the existing studies only reveal one side of the story. Table 1 shows majority of available performance evaluation and conventional coexistence researches.

### 3 802.11ah and 802.15.4g Coexistence Behavior and Coexistence Issue

Before conducting coexistence performance evaluation, we briefly introduce the functional differences between 802.11ah and 802.15.4g, which affect the coexistence behavior of 802.11ah and 802.15.4g.

802.11ah defines OFDM PHY and uses the ED-CCA with a threshold of -75 dBm per MHz for coexistence control with other non-802.11 systems. 802.15.4g specifies MR-FSK, MR-OFDM and MR-O-QPSK PHYs and only addresses coexistence among devices using different 802.15.4g PHYs. 802.15.4g ED threshold is lower than -75 dBm, e.g., its ED threshold is in [-100 dBm, -78 dBm] for FSK PHY.

802.11ah channel width is in the unit of MHz, i.e., 1 MHz/2 MHz/4 MHz/8 MHz/16 MHz. However, 802.15.4g channel width is in the unit of kHz, i.e., 200 kHz/400 kHz/600 kHz/800 kHz/1200 kHz. 802.11ah data rate ranges from 150 kbps to 78 Mbps for even one spatial stream. On the other hand, 802.15.4g data rate ranges from 6.25 kbps to 800 kbps.

802.11ah CSMA/CA and 802.15.4g CSMA/CA are much different. 1) 802.11ah allows immediate channel access. 802.15.4g, however, requires backoff no matter how long channel has been idle. 2) 802.11ah backoff is much faster than 802.15.4g backoff due to much smaller parameters as shown in Table 2, where 802.15.4g parameters are for FSK PHY operating in 920 MHz band. 3) 802.11ah requires backoff suspension, i.e., 802.11ah device must perform CCA in each backoff slot and can decrease backoff counter only if the channel is idle. On the other hand, 802.15.4g has no backoff suspension. 802.15.4g device performs CCA after the backoff

procedure completes.

The ED threshold, channel width, data rate and first two CSMA/CA features are in favor of 802.11ah. However, the third CSMA/CA feature is in favor of 802.15.4g. Theoretically, an 802.11ah packet can be infinitely delayed, but an 802.15.4g packet has bounded delay.

Table 2: 802.11ah and 802.15.4g CSMA/CA Parameters

802.11ah Param.	Value	802.15.4g Param.	Value
CCA Time	40 $\mu$ s	phyCCADuration	160 $\mu$ s
Slot Time	52 $\mu$ s	UnitBackoffPeriod	1160 $\mu$ s
SIFS Time	160 $\mu$ s	AIFS Time	1000 $\mu$ s
DIFS Time	264 $\mu$ s	SIFS Time	1000 $\mu$ s

Based on forementioned functional differences, the purpose of 802.11ah and 802.15.4g coexistence simulation is to explore how network traffic and network size affect the coexistence behavior of 802.11ah and 802.15.4g as well as what are the critical coexistence issues to be addressed.

We use packet delivery rate and packet latency as metrics to evaluate the coexistence performance. The packet delivery rate is measured as the ratio of number of packets successfully delivered and total number of packets transmitted. The packet latency is measured as time difference from the time packet transmission process starts to the time the packet receiving is successfully confirmed. In other words, the packet latency is given by  $BackoffTime + DataTXTime + AckWaitingTime + AckRXTime$ . The simulation setup is described in section 5.

In Figs. 1 and 2, solid lines represent 802.11ah network performance and dash lines illustrate 802.15.4g network performance. In addition, 50-20-20 indicates 50 nodes for each network, 20 kbps offered load for 802.11ah network, 20 kbps offered load for 802.15.4g network, and so on.

Fig. 1 shows packet delivery rate of 802.11ah network and 802.15.4g network. We have following findings: 1) For all scenarios, 802.11ah network delivers near 100% of the packet, which indicates that network traffic and network size have less impact on 802.11ah packet delivery rate. 2) 802.11ah network traffic has impact on 802.15.4g packet delivery rate. 802.15.4g network packet delivery rate decreases as 802.11ah network traffic increases. 3) 802.15.4g network traffic affects more on its packet delivery rate. 802.15.4g network packet delivery rate decreases significantly as its network traffic doubles. 4) The network size has little effect on 802.15.4g network packet delivery rate.

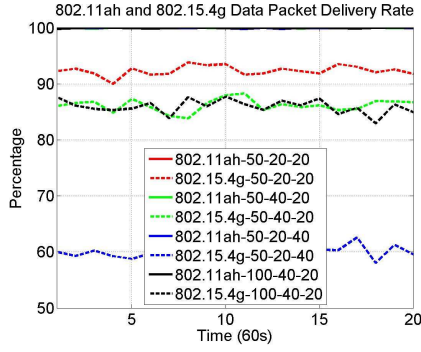


Figure 1: Packet Delivery Rate

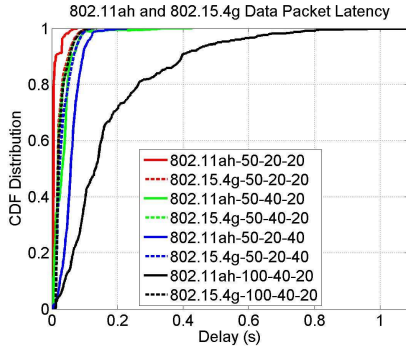


Figure 2: Packet Latency

Fig. 2 depicts the corresponding packet latency. We have following observations: 1) For all scenarios, 802.15.4g network achieves similar packet latency, which indicates that 802.15.4g packet is either delivered with the bounded delay or dropped and therefore, network traffic and network size have little impact on 802.15.4g packet latency. 2) 802.11ah network traffic has impact on its packet latency. 802.11ah packet latency increases as its network traffic increases. 3) 802.15.4g network traffic has more impact on 802.11ah packet latency. 802.11ah network packet latency increases more as 802.15.4g network traffic doubles. 4) Network size has major influence on 802.11ah packet latency. 802.11ah packet latency increases significantly as the number of nodes doubles, which verifies that 802.11ah packet can be infinitely delayed. These results show that 802.11ah network and 802.15.4g network interfere with each other. This observation is different from that drawn by existing studies that only reveal the 802.11ah interference on 802.15.4g. Based on these findings, coexistence technologies need to improve 802.15.4g delivery rate and reduce 802.11ah packet latency.

#### 4 Hybrid CSMA/CA for 802.15.4g to Coexist Better with 802.11ah

This section presents the proposed hybrid CSMA/CA for 802.15.4g to improve 802.15.4g delivery rate and reduce 802.11ah packet latency. The proposed hybrid CSMA/CA for 802.15.4g allows 802.15.4g device to perform immediate channel access.

An 802.15.4g device cannot communicate with an 802.11ah device. Therefore, 802.15.4g devices cannot coordinate with 802.11ah devices for interference mitigation

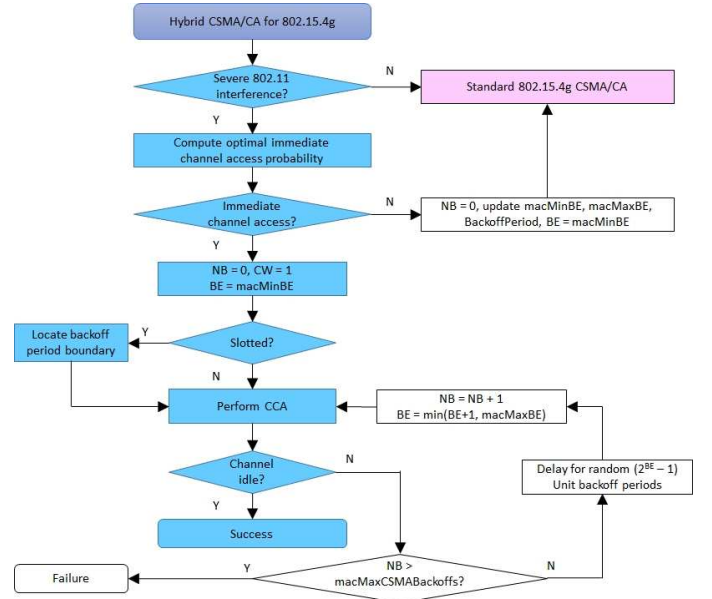


Figure 3: Hybrid CSMA/CA for IEEE 802.15.4g

without special assistance. However, 802.15.4g devices can explore the weakness of 802.11ah devices to increase their channel access opportunity when they detect severe interference from 802.11ah devices. An 802.11ah device must perform backoff process after the busy channel. Before the backoff process, 802.11ah device must wait for a DIFS ( $264 \mu s$ ) time period. This  $264 \mu s$  waiting time plus random backoff time gives 802.15.4g devices opportunity to start transmission before 802.11ah devices if 802.15.4g devices are allowed to have immediate channel access capability, which is not allowed in the 802.15.4g standard.

To compete with more aggressive 802.11ah for channel access, we propose an innovative hybrid CSMA/CA mechanism for 802.15.4g. Depending on severity of the 802.11ah interference, the hybrid CSMA/CA switches between two modes: immediate channel access disabled mode when 802.11ah interference is not severe and immediate channel access enabled mode when 802.11ah interference is severe. In the first mode, the standard 802.15.4g CSMA/CA is applied. In the second mode, the proposed immediate channel access enabled CSMA/CA is employed.

Fig. 3 shows the hybrid CSMA/CA mechanism. To decide a CSMA/CA mode, the hybrid CSMA/CA first determines the severity of 802.11ah interference. If the 802.11ah interference is not severe, the standard 802.15.4g CSMA/CA is applied. If the 802.11ah interference is severe, the immediate channel access enabled CSMA/CA is used. In this mode, the hybrid CSMA/CA enables 802.15.4g devices to have immediate channel access capability. The blue blocks show the flow chart of the immediate channel access. Consider that the immediate channel access by multiple 802.15.4g devices within a neighborhood may also cause collision, the hybrid CSMA/CA computes an optimal probability for stochastic decision making, i.e., perform immediate channel access or backoff.

To compute the optimal probability, an 802.15.4g

device first determines number of 802.15.4g neighbors by monitoring neighbor's packet transmission. Assume there are  $N_g$  802.15.4g devices in a neighborhood and each device has probability  $p$  to take immediate channel access and probability  $1 - p$  to perform backoff. Let  $X$  denote binomial random variable  $\sum_{i=1}^{N_g} X_i^g$ , where  $X_i^g (i = 1, 2, \dots, N_g)$  is random variable representing decision of 802.15.4g neighbor  $i$ . Then  $P(X = k) = \binom{N_g}{k} p^k (1 - p)^{N_g - k}$  and  $\mathbb{E}[X] = N_g p$ . To avoid collision among 802.15.4g transmissions due to immediate channel access, optimal strategy is that only one 802.15.4g device take immediate channel access and rest of 802.15.4g devices perform backoff, i.e.,  $\mathbb{E}[X] = 1$ , which gives optimal probability  $p_o = \frac{1}{N_g}$ .

Based on the optimal probability  $p_o$ , the hybrid CSMA/CA decides if immediate channel access or backoff is performed. The **Yes** decision leads to CCA operation. If the CCA returns idle channel, the immediate channel access takes place. The **No** decision leads to backoff. To do so, 802.15.4g device increases backoff parameters to avoid collision with transmission process of the immediate channel access device and also give 802.11ah device opportunity to transmit next and therefore, reduces 802.11ah packet latency.

## 5 802.11ah and 802.15.4g Coexistence Simulator

The existing simulation tools for 802.11 and 802.15.4, e.g., NS-3 [12] and OMNeT++, do not implement 802.11ah and 802.15.4g. Furthermore, to the best of our knowledge, there is no simulation tool that supports coexisting 802.11 and 802.15.4. We have developed an NS-3 based coexistence simulator for 802.11ah and 802.15.4g, in which we adopt the third party 802.11ah module [13] and implement 802.15.4g FSK PHY in the 920 MHz band. The challenges include the interfacing independent 802.11ah module and 802.15.4g module and the received power conversion.

Fig. 4(A) shows the developed interface between 802.11ah module and 802.15.4g module, where two modules notify each other with their transmission via a TX Info message that contains device position, transmission duration, transmission power, frequency, bandwidth, antenna gain, etc. Upon receiving TX Info message from other party, 802.11ah device and 802.15.4g device first compute the corresponding RX power  $P_{rx4g}$  and  $P_{rxah}$ , respectively, as shown in Fig. 4(B), where same transmission power is assumed. In other words, 802.11ah device computes 802.15.4g received power  $P_{rx4g}$  as if it was an 802.15.4g device and 802.15.4g device computes 802.11ah received power  $P_{rxah}$  as if it was an 802.11ah device. Using the received power computed, 802.11ah device and 802.15.4g device compute interference power level from other party as

$$\begin{aligned} P_{int}^{4g} &= P_{rxah}[\text{dBm}] - 10 \log_{10}(CH_{ah}/CH_{4g})[\text{dBm}], \\ P_{int}^{ah} &= P_{rx4g}[\text{dBm}], \end{aligned} \quad (1)$$

where  $P_{int}^{ah}$  is interference power to 802.11ah from 802.15.4g transmission,  $P_{int}^{4g}$  is interference power to 802.15.4g from 802.11ah transmission,  $CH_{ah}$  and  $CH_{4g}$  represent the

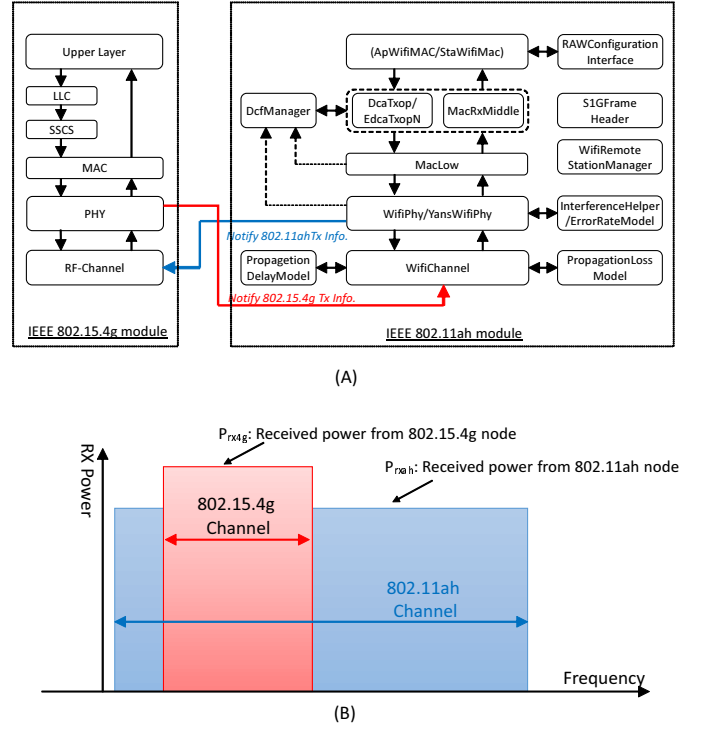


Figure 4: Sub-1 GHz Band Coexistence Simulator Model Interface

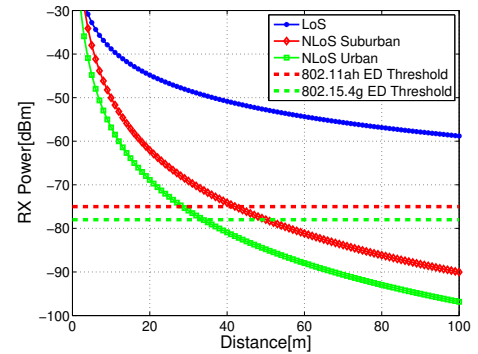


Figure 5: ITU-R P.1411-9 Propagation Model

channel width of 802.11ah channel and 802.15.4g channel, respectively. Using the interference power level and transmission duration, 802.11ah device and 802.15.4g device perform the enhanced CCA operation such that if the interference power is above the corresponding CCA-ED threshold, the channel status is considered as busy no matter what channel status is returned by their respective CCA operation.

Propagation model is another key component for practical simulation. NS-3 implements eight propagation models designed for general use scenarios without considering the emerging IoT applications. Both 802.11ah and 802.15.4g target the outdoor applications such as smart utility and smart city. Therefore, we adopt ITU-R P.1411-9 model for propagation between terminals located from below roof-top height to near street level. The median value of the

Non-Line-of-Sight (NLoS) loss is given by

$$L_{NLoS}^{median}(d) = 9.5 + 45 \log_{10} f + 40 \log_{10}(d/1000) + L_{urban}, \quad (2)$$

where  $f$  is the frequency,  $L_{urban}$  depends on the urban category and is 0 dB for suburban, 6.8 dB for urban, and  $d$  is the distance. Fig. 5 shows the propagation loss of LoS model, Suburban NLoS model and Urban NLoS model for transmission power of 13 dBm. With -78 dBm ED threshold, the intersection of the red curve and green dash line represents the effective energy detection distance for 802.15.4g, which is about 50 meters for Suburban NLoS model and 34 meters for Urban NLoS model. For 802.11ah with -75 dBm ED threshold, the corresponding distances are 42 meters and 28 meters, respectively.

## 6 Hybrid CSMA/CA Performance Evaluation

In this section, we evaluate the proposed CSMA/CA performance compared with standard 802.15.4g CSMA/CA. We adopt the simulation parameters recommended by IEEE 802.19 Working Group [14]. The frequency is in the 920 MHz band, transmission power is 13 dBm, 1 MHz channel for 802.11ah, 400 kHz channel for 802.15.4g, 802.11ah OFDM PHY rate is 300 kbps and 802.15.4g FSK PHY rate is 100 kbps. ITU-R P.1411-9 propagation model is employed in the simulation.

Typical two scenarios of [14] are simulated. One 802.15.4g network consists of 50 nodes uniformly deployed in a circle centered at PANC (Personal Area Network Coordinator) with radius of the effective energy detection distance. The PANC is located at (0, 0). Three 802.11ah networks are deployed inside 802.15.4g network with each 802.11ah network having 17 or 33 nodes uniformly distributed in a circle centered at corresponding AP with radius of the effective energy detection distance. Based on propagation model, three APs are located at (8, 0), (-4, 6.928), (-4, -6.928) and (6, 0), (-3, 5.196), (-3, -5.196), respectively. The offered network load is 20 kbps or 40 kbps. The offered network load is uniformly distributed among network nodes. The packet size is 100 bytes.

**Scenario-1:** The offered load for both networks is 20 kbps, i.e., 400 bps offered load per node, which leads to 0.13% duty cycle for 802.11ah node and 0.4% duty cycle for 802.15.4g node. These duty cycles are much lower than the 10% duty cycle specified in ARIB STD T108 standard [1]. With 100 bytes of packet size, each node generates 0.5 packet per second. For both standard CSMA/CA and hybrid CSMA/CA, Fig. 6 shows that 802.11ah network delivers 100% of the packet. The standard CSMA/CA delivers 92.37% of 802.15.4g packet. The hybrid CSMA/CA delivers 95.77% of 802.15.4g packet, i.e., 3.4% improvement without degrading 802.11ah packet delivery.

Fig. 7 shows that for both 802.11ah and 802.15.4g, standard CSMA/CA achieves shorter packet latency than the hybrid CSMA/CA due to less 802.15.4g packet delivered. 802.11ah has shorter packet latency than 802.15.4g. In this case, the hybrid CSMA/CA increases 802.11ah packet latency slightly.

Table 3: Packet Delivery Rate Comparison

	11ah		15.4g		
	Standard	Hybrid	Standard	Hybrid	Diff.
Scenario 1	100 %	100 %	92.4 %	95.8 %	3.4 %
Scenario 2	100 %	100 %	86.2 %	90.7 %	4.5 %

**Scenario-2:** The offered load is 40 kbps for 802.11ah network and 20 kbps for 802.15.4g network, i.e., the offered load is 800 bps for 802.11ah node and 400 bps for 802.15.4g node, which leads to 0.26% duty cycle and 0.4% duty cycle, respectively. These duty cycles are much lower than the 10% duty cycle limit. Each 802.11ah node generates 1 packet per second and each 802.15.4g node generates 0.5 packet per second. Fig. 8 shows that both standard CSMA/CA and hybrid CSMA/CA deliver near 100% of 802.11ah packet. The hybrid CSMA/CA improves 802.15.4g packet delivery rate from 86.2% given by standard CSMA/CA to 90.7%. This 4.5% improvement is done without degrading 802.11ah packet delivery. It indicates that as 802.11ah network traffic increases, the hybrid CSMA/CA provides more improvement on 802.15.4g packet delivery rate. Fig. 9 shows that 802.11ah and 802.15.4g have similar packet latency. For 802.15.4g, standard CSMA/CA achieves slightly shorter packet latency than the hybrid CSMA/CA due to less 802.15.4g packet delivered. However, the hybrid CSMA/CA maintain overall 802.11ah packet latency compared with [4].

Table 3 shows Packet Delivery Rate of 802.11ah and 802.15.4g for both standard 802.15.4g CSMA/CA and the proposed hybrid CSMA/CA for 802.15.4g. The hybrid CSMA/CA can improve 802.15.4g packet delivery rate by 4.5 % without degrading 802.11ah packet delivery rate in Scenario 2.

## 7 Conclusion

The heterogeneous wireless technologies developed for IoT applications increase the coexistence potential and present coexistence challenges. This paper takes IEEE 802.11ah and IEEE 802.15.4g as target technologies to investigate the Sub-1 GHz band coexistence. We evaluated 802.11ah and 802.15.4g coexistence behavior and identified 802.15.4g packet delivery rate and 802.11ah packet latency as the coexistence issues to be addressed. Accordingly, we proposed a hybrid CSMA/CA mechanism for 802.15.4g to achieve better coexistence with 802.11ah. To contend for channel access with more aggressive 802.11ah, the hybrid CSMA/CA allows 802.15.4g to perform immediate channel access. Using the developed Sub-1 GHz band coexistence simulator, we conducted the performance analysis of the proposed hybrid CSMA/CA. Compared with the standard 802.15.4g CSMA/CA, simulation results show that the hybrid CSMA/CA can improve 802.15.4g packet delivery rate by 4.5 % without degrading 802.11ah packet delivery rate. The hybrid CSMA/CA also maintain overall 802.11ah packet latency compared with conventional work [4].



## REFERENCES

- [1] ARIB STD-T108, "920MHz-Band Telemeter, Telecontrol and Data Transmission Radio Equipment," <https://www.arib.or.jp/english/html/overview/doc/5-STD-T108v1.0-E1.pdf>, 2012.
- [2] Ben A. Rolfe, "PAR as approved by REVCOM Dec 2019," IEEE802.19-18-0093r0, [https://mentor.ieee.org/802.19/documents/2018/ IEEE 802.19, Dec 2018](https://mentor.ieee.org/802.19/documents/2018/IEEE%20802.19%2C%20Dec%202018).
- [3] J. Guo and P. Orlik, "Self-transmission control in IoT over heterogeneous wireless networks," 2017 Ninth International Conference on Ubiquitous and Future Networks, Milan, 2017, pp. 898-903.
- [4] Y. Liu, J. Guo, P. Orlik, Y. Nagai, K. Watanabe and T. Sumi, "Coexistence of 802.11ah and 802.15.4g networks," 2018 IEEE Wireless Communications and Networking Conference, Barcelona, 2018, pp. 1-6.
- [5] W. Yuan, J. M. G. Linnartz and I. G. M. M. Niemegeers, "Adaptive CCA for IEEE 802.15.4 Wireless Sensor Networks to Mitigate Interference," 2010 IEEE Wireless Communication and Networking Conference, Sydney, NSW, 2010, pp. 1-5.
- [6] E. D. N. Ndihi and S. Cherkaoui, "Adaptive 802.15.4 backoff procedure to survive coexistence with 802.11 in extreme conditions," 2016 13th IEEE Annual Consumer Communications and Networking Conference, Las Vegas, NV, 2016, pp. 556-561.
- [7] X. Zhang and K. G. Shin, "Enabling Coexistence of Heterogeneous Wireless Systems: Case for ZigBee and WiFi," in Proceedings of the 12 th ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2011.
- [8] J. Hou, B. Chang, D.-K. Cho, and M. Gerla, "Minimizing 802.11 Interference on Zigbee Medical Sensors," in Proceedings of the Fourth International Conference on Body Area Networks, ICST, 2009.
- [9] J. W. Chong, C. H. Cho, H. Y. Hwang, and D. K. Sung, "An Adaptive WLAN Interference Mitigation Scheme for ZigBee Sensor Networks," in International Journal of Distributed Sensor Networks, 2015.
- [10] B. B. Olyaei, J. Pirskanen, O. Raeesi, A. Hazmi, and M. Valkama, "Performance Comparison Between Slotted IEEE 802.15.4 and IEEE802.11ah in IoT Based Applications," in 1st International Workshop on Internet of Things Communications and Technologies, IEEE, 2013, pp. 332-337.
- [11] R. Ma, S. Chen, H. Chen and W. Meng, "Coexistence of Smart Utility Networks and WLANs in Smart Grid Systems," in IEEE Transactions on Wireless Communications, vol. 15, no. 12, pp. 8313-8324, Dec. 2016.
- [12] The ns-3 network simulator, <http://www.nsnam.org/>.
- [13] L. Tian, L. Latre, and J. Famaey, "Implementation and validation of an IEEE 802.11ah module for NS-3," in Workshop in NS3, ACM, 2016.
- [14] Y. Nagai, J. Guo, T. Sumi, P. Orlik, and H. Mineno, "S1G Coexistence Simulation Profile," IEEE802.19-19-0021r2, [https://mentor.ieee.org/802.19/documents/2019/ IEEE 802.19, May 2019](https://mentor.ieee.org/802.19/documents/2019/IEEE%20802.19%2C%20May%202019).

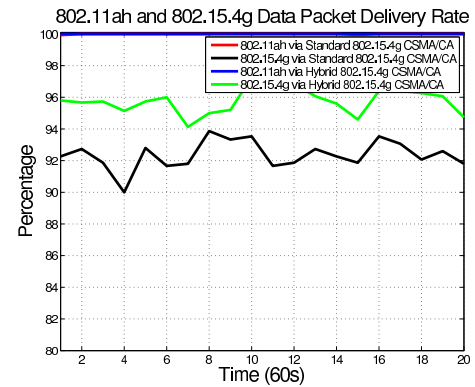


Figure 6: Packet Delivery Rate (Scenario 1)

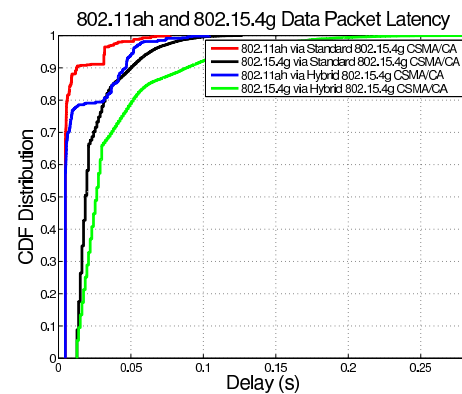


Figure 7: Packet Latency (Scenario 1)

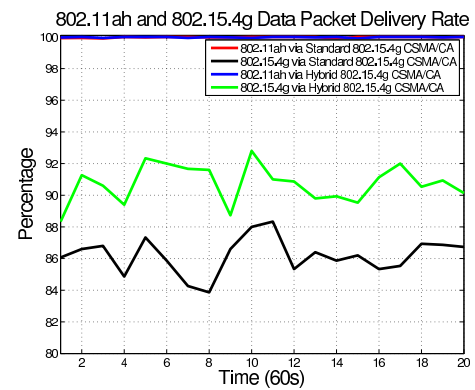


Figure 8: Packet Delivery Rate (Scenario 2)

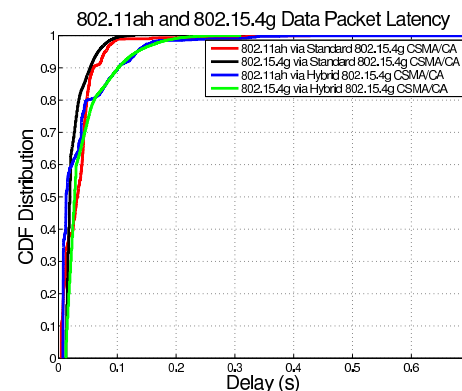


Figure 9: Packet Latency (Scenario 2)



Session 5:

ITS

( Chair: Tomoya Kitani )



# A Route Guidance Method for Vehicles to Reduce Traffic Congestion Considering Delay in Destination Arrival Time

Yusuke Matsui<sup>†</sup> and Takuya Yoshihiro<sup>‡</sup>

<sup>†</sup>Graduate School of Systems Engineering, Wakayama University, Japan

<sup>‡</sup>Faculty of Systems Engineering, Wakayama University, Japan

<sup>†</sup>e-mail : s216260@wakayama-u.ac.jp

<sup>‡</sup>e-mail : tac@wakayama-u.ac.jp

**Abstract** - In recent years, with the increase of the city population and the number of vehicles owned, traffic congestion has occurred in large cities. It has become a social problem. If traffic congestion can be alleviated, it will be effective for improving economic loss and environmental pollution. In related work, it is proposed to provide route guidance for vehicles on near congested roads. However, providing route guidance to only near the congested roads easily meets capacity limitation of traffic volume on the detour paths. It leads to congestion at the detour paths and delay in arrival time to destinations. In addition, the time required for all vehicles to reach their destinations cannot be minimized because the increased time due to detouring is not taken into consideration. In this paper, we propose a method to reduce the average travel time of all vehicles by considering the delay in arrival time caused by detouring.

**Keywords:** vehicular congestion avoidance, Intelligent Transportation System, Dynamic Route Guidance System, SUMO

## 1 Introduction

In recent years, with the increase of the city population and the number of vehicles owned, traffic congestion has occurred in large cities. It has become a social problem. According to study [1], the annual traffic congestion loss in Japan is about 3.81 billion person-hours, which is about 12 trillion yen when converted into monetary value. It also causes environmental problems. Also, the problem caused by traffic congestion is considered in the world. For example, in PM2.5 air pollution in Beijing, China, PM2.5 particles caused by vehicle exhaust gas accounts for about 20% [2]. By eliminating traffic congestion, PM2.5 emissions from vehicles were significantly reduced. Therefore, it is necessary to implement effective countermeasures for traffic congestion at major traffic congestion locations in cities.

On the other hand, from the driver's point of view, delay in arrival time due to traffic congestion is a big problem. If the traveling time of the vehicle can be reduced by alleviating the traffic congestion, their lost time can be minimized. As a result, the economic efficiency of the whole society will be improved.

Currently, a method called DRGS (Dynamic Route Guidance System) [3] is attracting attention as a means for reducing their negative effects. In DRGS, with RSU (Road Side Unit), which is communication terminals installed on the side of the road, we can monitor traffic conditions and prompt each vehicle to change its route. In this kind of studies, the authors have proposed route guidance systems that change the route of vehicles around the congested road to reduce the traffic jam. However, there is a problem that the invoked delay of vehicles in arrival time at the destinations is large because the offered routes largely change traveling time. Besides, even when the driver of each vehicle is advised to change the route, they would not possibly change their route if the offered route is not good and seems significantly to delay the arrival time. In constant, if the location of a vehicle is relatively far from the congested road, there will be an alternative route with smaller loss in time with higher probability. If we make route guidance that prioritizes the arrival delay, we would achieve congestion resolution plan that meets the minimum-delay requirement of each driver.

In this paper, we assume that RSU are installed on all the main roads and intersections in the target area so that the traffic conditions can be grasped completely. When a traffic jam occurs, a route guidance plan that tries to minimize the average arrival delay of each vehicle is calculated. Each RSU detects a congested road from vehicle density and speed in the road, and the server calculates the a feasible detour plan to eliminate traffic jam. By controlling the traffic with the rerouting plan, arrival delay of each vehicle is significantly reduced and the traffic congestion is eliminated by appropriately reducing in-coming vehicles to the congested road.

This work is organized as follows. In section 2, the outline of DRGS, its related work, and its problems are described. In section 3, we describe the proposed method. Finally in section 4, we summarize our study and describe future work.

## 2 Related Works

DRGS is a kind of systems that obtains road traffic information in real time and provides advised routes to vehicles. The purpose of DRGS is to reduce traveling time, reduce traffic congestion by providing the advised routes for vehicles.

Souza et al. [4] proposed a DRGS called CHIMERA (Congestion avoidance through a traffic classification MEchanism and a Re-routing Algorithm), which improved the overall spatial utilization of a road network and also reduced the average vehicle travel costs by avoiding vehicles from getting stuck in traffic. It is assumed that all vehicles are providing their information (vehicleID, current position, route, and destination) to a central entity through a single hop, long range communication such as 4G and LTE. CHIMERA is two main phases. The first one is congestion detection and traffic classification. The second one is route suggestion. The congestion detection and traffic classification processes depend on the traffic condition of all road segments. Traffic congestion is detected from the classification results of traffic condition using a k-Nearest Neighbor (kNN) classifier. In route suggestion process, CHIMERA computes alternative routes for all the vehicles based on probabilistic k-Shortest Paths. By distributing vehicles to multiple routes according to the probabilistic distribution, the possibility of generating new traffic congestion is reduced.

Pan et al. [5] proposed a centralized system to obtain the vehicle speed and density in real time in order to detect traffic jams. Once detected, vehicles are rerouted based on rerouting strategies such as DSP (Dynamic Shortest Path), RkSP (Random k Shortest Path), EBkSP (Entropy Balanced k-Shortest Paths) and FBkSP (Flow Balanced k Shortest Path). DSP is the classical rerouting strategy, which computes the routes with the smallest travel time. However, a shortcoming of this algorithm is that it may cause new congestion in another spot. RkSP randomly chooses a route among k shortest path routes to balance traffic. This strategy reduces the possibility of creating congestion in another spot. However, in RkSP, the routes are randomly selected so that reducing the travel time of all vehicles is not cared much about. EBkSP is a better strategy than RkSP. It adopts the smarter route selection, which considers the impact of each path selection on the future based on the idea of entropy. FBkSP computes k-shortest paths for route guidance and calculates the traffic volume of all road segments on each route. When providing a route to each vehicle, the route is chosen so that the traffic volume of each road segment in the area is well-balanced. As a result, the possibility of new traffic congestion is reduced.

The above methods [4] [5] would be the representative conventional methods for solving the problems we are targeting on, and have been selected as comparison methods in many related papers. However, these do not directly take the arrival delay optimization of each vehicle into account. Although the strategy detouring vehicles close to the congested point may reduce the time to resolve the traffic jam, it tends to increase the total traveling delay of vehicles. Furthermore, if only the vehicles around the congested road is detoured, the traffic volume on neighboring roads would easily be exceeded to lead another congestion. By using our strategy, it is possible to care for minimizing the impact of congestion on the entire

road traffic.

Shen et al. [6] proposed DRGS called NRR (Next Road Rerouting) that considers the feasibility of the system. It focused on reducing computational costs and system implementation costs. NRR is assumed to be deployed as a software plug-in for SCATS [7], which has already been in use at more than 37,000 intersections in 27 countries. Extending an existing system would considerably reduce the introduction cost. When NRR detects a congested road, it guides the vehicles around it. Specifically, instead of calculating the detour routes to the destinations of vehicles, NRR just calculates the alternative road segment that each vehicle should go next regardless of their destinations. After following the instruction, each vehicle travels to the destination via the shortest route based on its own VNS (Vehicular Navigation System) [8]. Since the functions in route guidance is limited, the calculation cost of the entire system can be reduced. However, this method has the similar problem to [4] [5] because the vehicles at close to the congested road are detoured.

### 3 Proposed Method

Due to traffic congestion, the time for each vehicle to reach its destination is significantly increased. Our purpose is to eliminate traffic congestion while minimizing the arrival delay of all cars. In related work, route guidance is applied to vehicles located close to the congested road. Although the strategy detouring vehicles close to the congested point may reduce the time to resolve the traffic jam, it tends to increase the total traveling delay of vehicles. Our method calculates how much the arrival delay to the destination is when a vehicle is guided to bypass a congested road at intersection. Then it guides the vehicle to the detour path with the smaller arrival delay. As a result, it is possible to eliminate the traffic congestion while minimizing the effect on the arrival delay of each vehicle to the destination.

The details of our method are described as follows. In section 3.1, the notation used in our method is described. In section 3.2, we describe the method to construct the rerouting table. The rerouting table manages detour routes to guide vehicles and enables us to judge which detour routes be activated according to the magnitude of the occurring traffic congestion. In section 3.3, we explain the algorithm for controlling the vehicle based on the rerouting table.

#### 3.1 Notation

A directed and weighted graph  $G = (I, R)$  represents the road network, where  $I$  is the set of intersections,  $R$  is the set of road segments,  $S \subset I$  is the set of source points and  $D \subset I$  is the set of destinations of vehicles.  $w(\cdot)$  is a function that represents the weight of the road segment:  $w(r)$  is a positive integer and represents the weight for road segment  $r \in R$ . For a source  $s \in S$  and destination  $d \in D$ ,  $P_{(s,d)}$  represents

the shortest path from  $s$  to  $d$  under the weight function  $w(\cdot)$ .  $B_{(s,d)}$  represents the detour path from  $s$  to  $d$  defined as the shortest path under the weighting function  $w(\cdot)$  in the road network  $G' = (I, R - P_{(s,d)})$ , where  $P_{(s,d)}$  also represents the set of road segments included in  $P_{(s,d)}$ . Namely,  $B_{(s,d)}$  is the shortest path from  $s$  to  $d$  in  $G$  that is the edge-disjoint to  $P_{(s,d)}$ . For route  $P$ ,  $w(P)$  is the weight of  $P$  defined as the sum of the weights of all road segments included in  $P$ , i.e.,  $w(P) = \sum_{r \in P} w(r)$ .  $C_{(s,d)}^{diff}$  represents the difference in weight between the shortest path and the detour path from  $s$  to  $d$ , where  $C_{(s,d)}^{diff} = w(B_{(s,d)}) - w(P_{(s,d)})$ .

$V$  is a set of vehicles existing in the road network.  $V_{len}^{avg}$  is the average vehicle length, and  $g_{min}$  is the minimum inter-vehicular distance. For road segment  $r$ ,  $len_r$  is the segment length and  $lane_r$  is the number of lanes,  $n_r^t$  is the number of vehicles on road  $r$  at time  $t$ . The vehicle density of  $r$  at time  $t$  is expressed as follows

$$K_r^t = \frac{n_r^t}{len_r}. \quad (1)$$

The maximum number of vehicles that can exist on  $r$  is  $n_r^{max} = \frac{len_r}{V_{len}^{avg} + g_{min}} \times lane_r$ . The maximum vehicle density of  $r$  is given as

$$K_r^{max} = \frac{n_r^{max}}{len_r}. \quad (2)$$

At time  $t$ , when the ratio of vehicle density to the maximum vehicle density of road  $r$  exceeds a predefined threshold  $\delta$ , i.e.,  $\frac{K_r^t}{K_r^{max}} \geq \delta$ ,  $r$  is regarded as congested road.

The traffic volume on road  $r$  is defined as the number of vehicles passing  $r$  per unit time as follows

$$F_r^t = \nu_r^t \times K_r^t, \quad (3)$$

where  $\nu_r^t$  represents the average speed of vehicles on road segment  $r$  at time  $t$ .

In this study, the state where no traffic congestion occurs is called the steady-state.  $\nu_r^{std}$  represents the vehicle speed on the road segment  $r$  in the steady-state. The traffic capacity of road segment  $r$  is defined as

$$F_r^{cap} = \nu_r^{std} \times \delta K_r^t. \quad (4)$$

$F_r^{cap} - F_r^t$  is called the allowable traffic volume of  $r$ .

Assume that, at time  $t$ , the shortest path  $P_{(s,d)}$  from intersection  $s \in S$  to  $d \in D$  passes through road segment  $r$ , and the congestion is detected at  $r$ . If we consider to guide vehicles that travels from  $s$  to  $d$  to bypass on  $B_{(s,d)}$ , the allowable traffic volume for  $B_{(s,d)}$  is defined as

$$A_{(s,d)}^t = \min_{r \in B_{(s,d)}} (F_r^{cap} - F_r^t). \quad (5)$$

Therefore, we can guide the vehicles heading for the intersection  $d$  at the intersection  $s$  to the detour route  $B_{(s,d)}$  up to the upper limit  $A_{(s,d)}^t$ .

The notations used above are shown in the Table 1.

Table 1: Notation

Symbol	Description
$G$	graph (road topology)
$I$	intersection set
$R$	road segment set
$S$	source points set
$D$	destinations set
$w(r)$	weight of road segment $r \in R$
$w(P)$	weight of route $P$
$P_{(s,d)}$	shortest path from intersection $s \in S$ to intersection $d \in D$
$B_{(s,d)}$	detour path from intersection $s \in S$ to intersection $d \in D$
$C_{(s,d)}^{diff}$	the difference of weight between the shortest path and detour path from $s$ to $d$
$V$	vehicle set
$V_{len}^{avg}$	average of vehicle length
$g_{min}$	minimum inter-vehicular distance
$K_r^t$	vehicle density of $r$ at time $t$
$n_r^t$	vehicle number of $r$ at time $t$
$\nu_r^t$	average speed of the vehicles run on $r$ at time $t$
$K_r^{max}$	maximum vehicle density of $r$
$n_r^{max}$	maximum vehicle number of $r$
$\nu_r^{std}$	average vehicle speed of $r$ in steady-state
$len_r$	road length of $r$
$lane_r$	lane number of $r$
$\delta$	threshold at congestion detection
$F_r^t$	traffic volume of $r$ at time $t$
$F_r^{cap}$	traffic capacity of $r$
$A_{(s,d)}^t$	allowable traffic volume of $B_{(s,d)}$

## 3.2 Reroute Planing

### 3.2.1 Grasping Traffic Volume In Steady State

The digital road map of the target area is obtained in advance, i.e., we can obtain not only  $G = (I, R)$  but also the road length  $len_r$  and the number of lanes  $lane_r$  on each road  $r \in R$ . It is assumed that the traffic condition is always measured by RSU installed beside the road, i.e., RSUs can obtain the average vehicle speed  $\nu_r^t$  and the number of vehicles  $n_r^t$  on each road segment  $r$  at any time  $t$ . It is also assumed that each vehicle has original route given by its own VNS and the route is transmitted to the server through RSUs. Therefore, it is possible to calculate  $T_{(s,d)}^t$  from vehicle route, where  $T_{(s,d)}^t$  is the number of vehicles arriving at the intersection  $s$  to destination  $d$  per unit time at time  $t$ . The weight  $w(r)$  of each road  $r$  is represented by the time required to pass the road  $r$ :  $w(r) = \frac{len_r}{\nu_r^t}$ . We can calculate the average vehicle speed  $\nu_r^{std}$  and the number of vehicle  $T_{(s,d)}^{std}$  in steady-state with the above values.

### 3.2.2 Rerouting Table

The rerouting table essentially represents the priority of intersections to apply route guidance, and is calculated based on the steady-state traffic volume when traffic congestion on road segment  $r$  is detected. As shown in Table 2, rerouting table consists of priority  $L_{(s,d)}$ , source intersections  $s \in S$ , destination intersections  $d \in D$ , arrival delay  $C_{(s,d)}^{diff}$ , traffic volume to guide  $E_{(s,d)}$  and  $X_{(s,d)}$ , where  $X_{(s,d)}$  is the cumulative value of  $E_{(s,d)}$ . The priority  $L_{(s,d)}$  is a positive integer sequentially assigned in the ascending order of  $C_{(s,d)}^{diff}$ . Here, for a pair of  $s \in S$  and  $d \in D$ , the intersection pair is represented by  $u = (s, d)$ , i.e., the priority is shown as  $L_u (= L_{(s,d)})$ , the traffic volume to guide as  $E_u (= E_{(s,d)})$ , and the cumulative value of  $E_u$  as  $X_u (= X_{(s,d)})$ .  $X_u$  is the sum of  $E_{u'}$  for all intersection pairs  $u'$  whose priority is  $L_u$  or less.

When congestion on road segment  $r$  is detected at time  $t$ , we reduce the traffic volume injecting to road  $r$  to eliminate the traffic jam on road  $r$ . The traffic volume required to reduce traffic jam is expressed as follows

$$F_r^{exc} = F_r^{std} - F_r^t. \quad (6)$$

Where  $F_r^{std}$  is the traffic volume on road  $r$  in steady-state,  $F_r^t$  is the traffic volume on road  $r$  at time  $t$ . We consider that the congestion will be resolved by guiding vehicles passing through the road  $r$  to the detour path by  $\beta F_r^{exc}$ , where  $\beta$  is the congestion expansion factor  $\beta \geq 1$ .

The pairs of intersections (i.e., guidances) applied to vehicles are selected in the ascending order of priority in the rerouting table in order to eliminate the congestion on road  $r$ . The method to select the pairs of intersection to apply from the rerouting table is as follows

$$\alpha X_u \geq \beta F_r^{exc}, \quad (7)$$

where  $\alpha$  is the probability of guided vehicles to follow route guidance. If the cumulative value of traffic that would be guided exceeds the traffic volume to be reduced, i.e.,  $\alpha X_u$  exceeds  $\beta F_r^{exc}$ , the congestion on road  $r$  will be resolved before long. If the minimum value of the priority that fulfills Equation(7) is  $j$  (remember that  $j$  determines  $X_u$ ), route guidance is applied at the intersections with the priority less than or equal to  $j$ .

For instance, assume that  $F_r^{exc} = 85$ ,  $\alpha = 0.7$ ,  $\beta = 1.3$ . If route guidance is applied at the pairs of intersection up to the sixth line in Table 2 (i.e.,  $k = 6$ ),  $\alpha X_u = 115.5$  vehicles per unit time can be guided to avoid the congested road  $r$ . Since the traffic volume to guide exceeds  $\beta F_r^{exc} = 110.5$ , the congestion will be eliminated by route guidance with these six intersection pairs.

### 3.2.3 Constructing Rerouting Tables

The rerouting table is constructed by the following procedure, where  $U^{all}$  is the set of all the intersection pairs in the target area.

- (1) The pair of intersections  $u = (s, d) \in U^{all}$  is extracted in the ascending order of  $C_u^{diff}$ .
- (2) If  $r \notin P_u$ , skip  $u$  and proceed to the next pair of intersections.
- (3) If there is an intersection pair  $u' = (x, d)$  in the rerouting table where  $P_u \subset P_{u'}$  or  $P_u \supset P_{u'}$ , skip  $u$  and proceed to the next intersection pair.
- (4) The allowable traffic volume  $A_u$  for the detour path in the steady state is set to the guiding traffic volume  $E_u$  (i.e.,  $E_u = A_u$ ), and add the corresponding entry to the rerouting table. ( $L$  is incremented by one every time an entry is added.)
- (5) We update  $A_{u'}$  for all  $u' \in U$  with the values assuming that the traffic volume  $E_u$  is bypassed for  $u$ .

In steps (1) and (2), all intersection pairs included in  $U^{all}$  are looped in the increasing order of  $C_u^{diff}$ .

Step (3) is a process to avoid guiding a single vehicle multiple times. Suppose two intersection pairs with the same destination  $u_1 = (s_1, d)$  and  $u_2 = (s_2, d)$ . Then, assume  $P_{u_1}$  is included in  $P_{u_2}$  (i.e.,  $P_{u_1} \subset P_{u_2}$ ). In this case, if  $C_{u_2}^{diff} < C_{u_1}^{diff}$ , the vehicles guided with  $u_1$  at  $s_1$  has already been guided with  $u_2$  at  $s_2$ . We consider that guiding a vehicle many times will be a burden for the driver, and so we avoid this. Therefore, in this case, route guidance for  $u_1$  is not applied. On the other hand, if  $C_{u_2}^{diff} > C_{u_1}^{diff}$ , the vehicle guided with  $u_2$  at  $s_2$  is again guided with  $u_1$  at  $s_1$ . Therefore, in this case, route guidance for  $u_2$  is not applied. As a result, the intersection pair  $u$  that has the same destination as the intersection pairs with the smallest  $C_{(s,d)}^{diff}$  is not included in the rerouting table.

In step (4), for each pair of intersections, the traffic volume to guide the detour path is determined based on the road capacity.

In step (5), since the traffic volume on each road changes as a result of route guidance, the allowable traffic volume for each intersection pair is updated. To update the allowable traffic volume in step (5), follow the procedure below.

- (a) For each road segment  $r' \in P_u$ , subtract  $E_u$  from  $F_{r'}^{std}$ .
- (b) For each road segment  $r' \in B_u$ , add  $E_u$  to  $F_{r'}^{std}$ .
- (c) Update  $A_{u'}$  for every intersection pair  $u' \in U^{all}$ .



Table 2: Rerouting Table

priority $L_{(s,d)}$	source intersection $s \in S$	destination intersection $d \in D$	arrival delay $C_{(s,d)}^{diff} [min]$	traffic volume guided $E_{(s,d)} [/min]$	cumulative volume $X_{(s,d)} [/min]$
1	$s_1$	$d_1$	3	26	26
2	$s_2$	$d_2$	3	28	54
3	$s_1$	$d_2$	6	25	79
4	$s_3$	$d_1$	8	21	100
5	$s_1$	$d_3$	8	34	134
6	$s_4$	$d_2$	9	31	165
7	$s_3$	$d_4$	11	14	179
:	:	:	:	:	:
:	:	:	:	:	:
k-1	$s_7$	$d_{12}$	31	$E_{(s_7,d_{12})}$	$X_{(s_7,d_{12})}$
k	$s_{14}$	$d_5$	33	$E_{(s_{14},d_5)}$	$X_{(s_{14},d_5)}$
:	:	:	:	:	:

### 3.3 Rerouting Algorithm

This section describes the procedure for providing routes to vehicles based on the rerouting table created in Section 3.2. First, if congestion at road segment  $r$  is detected at time  $t$ , the current excessing traffic volume of road  $r$  is calculated, i.e.,  $F_r^{exc} = F_r^{std} - F_r^t$ . Each vehicle has its own destination  $d \in D$ , and also its own shortest path. The source intersection  $s$  and the destination intersection  $d$  are acquired from every vehicle  $v$ . At this time, it is assumed that the RSU can communicate with the vehicle at each intersection. After finding the intersection pair  $u = (s, d)$  of each vehicle  $v$ , if the priority of  $u$  is equal to or less than  $j$  ( $j$  is the minimum value of the priority that fulfills Equation(7)), the detour path  $B_u$  is provided to  $v$  for route guidance. The vehicles guided to the detour path are expected to change the route with probability  $\alpha$ , and does not change with the probability  $1 - \alpha$ .

## 4 Conclusion

In this paper, we proposed a route guidance method that are aware of the arrival delay to the destinations when vehicles bypass a congested road. Specifically, for each intersection, we care the difference in travel time between the shortest path and the detour path. The intersection pair with the smaller deference is given the priority in applying route guidance. This enables us to reduce traffic congestion while suppressing the arrival delay to the destinations. As future work, we would evaluate that the proposed method can reduce the average travel time compared to conventional methods, and alleviate the traffic congestion. Also we would clarify that the proposed method reduces the average travel time of all vehicles that follows route guidance.

## REFERENCES

- [1] Ministry of Land, Infrastructure and Transport HP, <https://www.mlit.go.jp/road/index.html>, 2019-01-01
- [2] Wang, J. , Hu, M. , Xu, C. , Christakos, G. and Zhao, Y. "Estimation of Citywide Air Pollution in Beijing", (PLoS ONE, Volume: 8, No.1, e53400, Jan. 2013)
- [3] J.M. Sparmann. "Benefits of dynamic route guidance systems as part of a future oriented city traffic management system", Vehicle Navigation and Information Systems Conference, (Oct. 1991)
- [4] Allan M. De Souza, Roberto S.Yokoyama, Guilherme Maia, Antonio Loureiro and Leandro Villas. "Real-Time Path Planning to Prevent Traffic Jam Through an Intelligent Transportation System ", IEEE Symposium on Computers and Communication (Volume : 1, pages 726-731, August. 2016)
- [5] J. Pan, I. Sandu Popa, K. Zeitouni, and C. Borcea. "Proactive vehicular traffic re-routing for lower travel time" , IEEE Transactions on Vehicular Technology (Volume : 62 , Issue : 8 , Oct. 2013)
- [6] Shen Wang, Soufiene Djahel, Zonghua Zhang and Jennifer McManis. "Next Road Rerouting(NRR) : A Multi-Agent System for Mitigating Unexpected Urban Traffic Congestion ", IEEE Transactions on Intelligent Transportation Systems (Volume: 17, Issue: 10 , Oct. 2016)
- [7] A. G. Sims, K. W. Dobinson. "The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits", IEEE Transactions on Vehicular Technology (Volume: 29, Issue: 2, May. 1980)
- [8] Sadayuki Tsugawa, Masayoshi Aoki, Akio Hosaka, and Kaoru Seki. "A survey of present ivhs activities in Japan " (Volume : 5, pp1591-1597, Nov. 1997)



# Examination of Environment Recognition Method for Autonomous Driving on Undeveloped Road

Kazumasa Kamitani<sup>†</sup> and Naoya Chujo<sup>‡</sup>

<sup>†</sup>Graduate School, Aichi Institute of Technology, Japan

<sup>‡</sup>Aichi Institute of Technology, Japan  
{b19710bb, ny-chujo}@aitech.ac.jp

**Abstract** - Autonomous driving technology has been widely adopted to reduce automobile accidents and to cope with adverse environmental conditions. However, autonomous driving technology has yet to be perfected on private and undeveloped roads. Conducting autonomous driving on public roads near residential areas can be complicated when the road environment is undeveloped or not maintained. In such environments, vehicle control can be complicated by factors such as gravel surfaces and potholes. In this study, we examined obstacle recognition on undeveloped road environments by an autonomous driving system. Specifically, we developed a method for estimating the spatial characteristics of gravel surfaces and the locations of potholes using LiDAR, and evaluated and verified the method in a virtual driving environment.

**Keywords:** Autonomous driving, undeveloped road, virtual environment, obstacle recognition, road environment

## 1 INTRODUCTION

In recent years, autonomous driving technology has become increasingly widespread as a means of reducing vehicle accidents and for handling adverse environmental conditions. For example, Mercedes-Benz[1] in Germany and Tesla[2] in the U.S. have tested autonomous vehicles on highways and trunk roads. However, study on autonomous driving technology have not been undertaken in areas, such as residential areas and private land, where the asphalt roads are undeveloped [3]. Hence autonomous driving system is not complete in all environments. In this study, we refer to maintained surfaces as flat roads with no potholes or humps, such as a highway. Undeveloped roads are roads that may be surfaced with asphalt, but they may have cracks, potholes, and humps, or they may simply be surfaced with gravel. When using an autonomous driving system, it is not easy to transition from maintained public roads to undeveloped driveways and parking lots, as the road environment in these areas can be unpredictable. However, an autonomous driving system must make appropriate decisions and operate automatically, even on undeveloped roads.

One of the solutions to this problem is to develop an autonomous driving system that uses remote driving[4], which is what we are currently working on. By driving remotely on roads where autonomous driving is not currently possible, we can obtain point cloud maps and the driving routes that will be required by autonomous driving systems to operate effectively in the future. However, we believe that vehicle control in underdeveloped environments is potentially hazardous be-

cause it interferes with the safe operation of the vehicle.

To resolve this problem, we addressed the perception of the road environment by an autonomous obstacle recognition system on an undeveloped road. On an undeveloped road, controlling a vehicle can be difficult due to obstacles such as humps and potholes. A hump is defined here as an elevation in the road with a parabolic shape that has a length of 3.6 m, a width of 3 m, and a height of about 10 cm. The existence of humps on roads can restrict the speed of a vehicle[5]. Therefore, if the vehicle is driven over a hump without any deceleration, the vehicle can become unstable and dangerous.

Potholes can adversely affect vehicle control. A pothole is defined here as a depression in the asphalt pavement surface with a diameter of 0.1 to 1 m [6]. Potholes have been reported to cause car accidents[7]. If the vehicle cannot decelerate or evade the pothole, the vehicle can become unstable and dangerous. Hence, we believe that identifying obstacles in the road environment will facilitate vehicle control and improve the functioning of the autonomous driving system.

This study uses Light Detection And Ranging (LiDAR) to identify features in the road environment. Autonomous vehicles are equipped with various sensors such as cameras, LiDAR, radar, GPS, and wheel encoders. LiDAR plays a particularly important role in recognizing the road surface and the surrounding environment.

Potholes and humps recognition is also researched in the reference[8]. The sensor that identified the hump is a LiDAR with a single laser. There are more than 16 lasers in LiDAR for autonomous driving vehicles. Therefore, we believe that better results than those in the literature can be obtained by using LiDAR in self-driving cars. Also, in the reference, the pothole uses a camera. However, it is difficult to obtain depth information with the camera. Therefore, we believe that the depth can be identified by using LiDAR.

We conducted an experiment using a virtual environment that considers the aforementioned problems. By using a virtual environment, it is easy to create a road environment that is undeveloped. In addition, in a virtual environment, experiments can be conducted without the risk of accidents.

The remainder of this manuscript is organized as follows. Section 2 describes the methods and proposed methods used to identify potholes and humps on maintained surfaces and gravel surfaces. Section 3 describes the experimental methods and results to verify whether the proposed method was effective. Section 4 discusses the experimental results, and Section 5 concludes the paper.

## 2 PROPOSED SYSTEM

In this section, an autonomous driving system with road obstacle-recognition capabilities is described, as shown in Fig. 1.

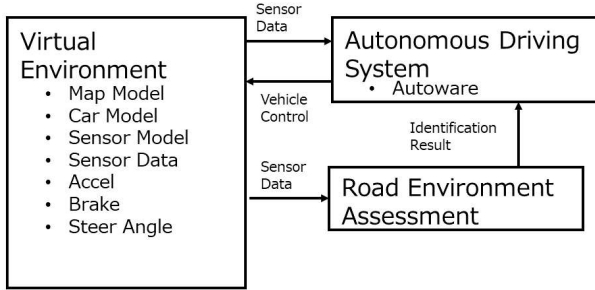


Figure 1: Proposal of road surface environment identification method

The virtual environment used in this experiment was the Unity system[9]. We introduced vehicle models, sensor models and map models using the LGSVL Simulator[10]. The autonomous driving system uses Autoware[11], an autonomous driving software platform. Autoware receives sensor information from the virtual environment and can estimate the location of an autonomous vehicle based on that information, and can identify objects around the vehicle.

It is possible to identify objects around vehicles, such as people and obstacles, but the road environment is identified using point cloud maps[12]. They created a point cloud map and then determined the unevenness of the road surface. The point cloud map is created after driving a LiDAR-equipped vehicle on the road being tested. Therefore, it is not possible to identify the road environment while driving. Since potholes may form at locations that are not reflected in the point cloud map, we considered it necessary to identify them when driving. We therefore clarified the features on the road environment based on sensor information obtained from the virtual environment, and determined the location of potholes and humps using LiDAR data. The results showed that it is necessary to plan a route and to slow down or avoid the obstacles. Recognition of the road environment was undertaken using Robot Operating System[13].

This study considered maintained surfaces and gravel-covered (i.e. undeveloped) surfaces, as well as obstacles such as potholes and humps as components of the road environment. First, the method used to clarify features of the road environment is explained.

### 2.1 Identification of road environment components

Obstacles, which varied along the road environment, included gravel surfaces, potholes, and humps. This section clarifies how features such as potholes and humps are identified on maintained surfaces and gravel surfaces.

#### 2.1.1 Maintained surface

A maintained surface is identified when three conditions are met.

1. Height data obtained by the LiDAR system is equal to the height of the LiDAR system from the ground
2. Point data in the forward direction of the vehicle is available
3. The Euclidean distance from the LiDAR system to the point data is less than 20 m

Firstly, in the case of a paved road, we assumed that the height of the installed LiDAR system is approximately equal to the height of the point data measured on the road. The error in the LiDAR data was  $\pm 5$  cm[14]. When the difference between the height data and the height of the LiDAR system from the ground was within the range of the LiDAR error, the point data were considered to be typical of a maintained surface.

Secondly, the range of obstacle recognition is limited to the front of the autonomous vehicle because of the assumed driving conditions. In this study, we employed a fan-shaped scanning area with an angle of  $120^\circ$ .

We established a third condition for two reasons. For the first reason, Fig. 2 shows an example of LiDAR point data. The yellow dots are the point data obtained by the LiDAR system, and the autonomous vehicle is at the center of the circles. LiDAR data further away from the vehicle are characterized by having a longer distance between laser channel(e.g., I and II in Fig. 2). As a result, recognizing a pothole with a diameter of less than 1 m using LiDAR data is difficult.

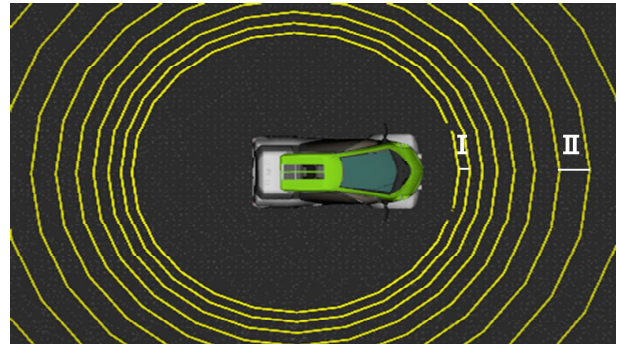


Figure 2: Example LiDAR data

For the second reason, recognizing a pothole with a diameter of less than 1 m is also complicated by the fact that as distance from the vehicle increases, the distance between the sensor and the vehicle becomes larger, making it more difficult to identify potholes. The evaluation criterion for stable vehicle control was ride quality. As shown in [15], while acceleration at  $0.9 \text{ m/s}^2$  was not unpleasant, that at  $1.2 \text{ m/s}^2$  was considered somewhat unpleasant. Since we assumed that the environment is a private road, and the traveling speed is 10 km/h, the stopping distance is 3.22 m. Therefore, if the feature recognition distance is at least 3.22 m, then ride quality is considered to be satisfactory.

### 2.1.2 Gravel surface

Gravel surfaces such as parking lots have many crushed stones with a maximum size of 2 cm [16]. Therefore, a gravel surface scanned by LiDAR comprises stones that are similar to those used for a maintained surface. To identify a gravel surface, we use the variance value with a height that is equivalent to that of a maintained surface. Let  $\sigma^2$  be the variance of the height data obtained for a maintained surface created in a virtual environment. Let the height of the point data for a maintained surface be  $M_z$ , let the total number of data points be  $TM_z$ , and let the average height be  $\overline{M_z}$ . We can then compute  $\sigma^2$  as

$$\sigma^2 = \frac{1}{TM_z} \sum_{k=1}^{TM_z} (M_{z_k} - \overline{M_z})^2 \quad (1)$$

As a result of the recognition of the paved and gravel surface, the maintained surface was designated  $5.81 \times 10^{-4} \text{ m}^2$  and the gravel surface was designated  $7.24 \times 10^{-4} \text{ m}^2$ . Based on these results, the threshold for distinguishing a gravel surface from a paved surface is  $6.4 \times 10^{-4} \text{ m}^2$ .

### 2.1.3 Pothole

A pothole is identified when the following three conditions are met.

1. Point data in the forward direction of the vehicle is available
2. When point data are obtained that are slightly larger than the height for a maintained surface
3. When the LiDAR system identifies two or more consecutive laser channels

For the first condition is the same as in sub section 2.1.1.

For the second condition, the depth of the pothole is below the road surface because it is a hole in the road. The threshold for the pothole depth at which the vehicle becomes unstable was set to 10 cm, based on [7].

For the third condition, in the case where a single laser channel registered a pothole, the possibility of an anomaly was considered. In addition, the size of the pothole cannot be estimated using one laser channel alone. We considered that small potholes can be ignored, and only those potholes that are problematic for vehicle control in automated vehicles were considered. Consequently, two or more laser channels were selected for this purpose.

### 2.1.4 Hump

A hump, such as that shown in Fig. 3, exists at the entrance of the parking lot on the AICHI INSTITUTE OF TECHNOLOGY campus, which is private property.

Therefore, we believe that the system needs to identify humps. Humps are identified when the following four conditions are met.

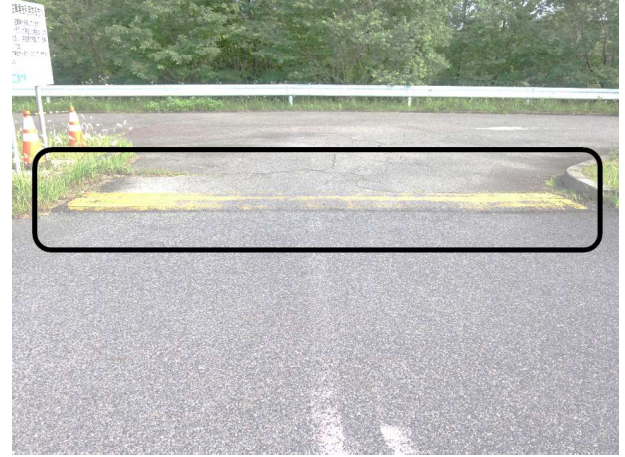


Figure 3: Hump in parking lot

1. Point data in the forward direction of the vehicle is available
2. When point data are obtained that are slightly less than the height of the maintained surface
3. When the LiDAR system identifies two or more consecutive laser channels
4. When the height registered by one laser pulse is higher than that for the subsequent laser channel (i.e. the channel closer to the vehicle) when the LiDAR continuously identifies point data higher than the maintained surface

For the first condition is the same as in sub section 2.1.1.

For the second condition, the hump is higher than the road surface, the height of humps was set at 10 cm, based on [5].

For the third condition, in the case of a single laser channel that identified a hump, the possibility of misidentification is considered and the size of the hump cannot be estimated. Therefore, two or more consecutive laser channels were selected for this purpose.

For the fourth condition, this condition is set not to identify a non-hump as a hump[8]. Humps are characterized by an obstacle increasing in height from the front to the back. The feature is identified as a hump when this condition is met.

## 3 EXPERIMENTAL EVALUATION

In order to verify whether the proposed method is capable of recognizing obstacles in the road environment effectively, we conducted an experiment in a virtual environment, which we describe in the following section.

### 3.1 Experimental virtual environment

The driving environment is shown in Fig. 4.

The area enclosed by the black rectangle in Fig. 4 is a parking lot with a gravel surface. The area enclosed by the blue rectangle is a section of road where hump is installed. The area enclosed by the red rectangle is a road with potholes. The green arrow indicates the direction of travel. The speed



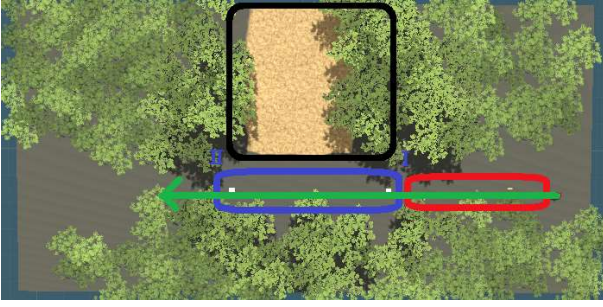


Figure 4: Experimental environment

of the autonomous vehicle was set to be 10 km/h, based on [17]. The dimensions of each hump are shown in Table 1.

Table 1: Hump data

	Length (m)	Width (m)	Height (m)
Hump I	3.7	3	0.075
Hump II	4.25	3	0.1

Fig. 5 shows an oblique perspective of the potholes enclosed by the red rectangles. A pothole is defined here as a depression in the asphalt pavement surface. Therefore, no pothole shall be allowed on the gravel road.

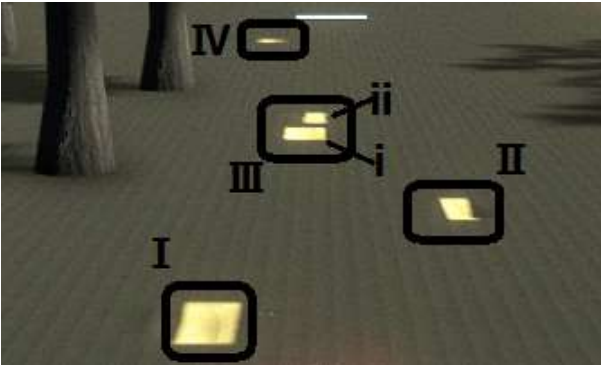


Figure 5: Example potholes

Pothole III is designed to be a multi-part pothole. The dimensions of each rectangular pothole are shown in Table 2.

The LiDAR system used in this experiment was assumed to be a VLP-32 system[14]. The performance of the LiDAR system is shown in Table 3. The vehicle used was a small electric vehicle. The height from the ground to the LiDAR platform was set to 1.75 m, considering the height of the vehicle. The distance from the LiDAR system to the front of the vehicle was set to 2.01 m. Also the development environment is shown in Table 4.

### 3.2 Experimental results

The results obtained for recognizing maintained surfaces, Gravel surfaces, potholes, and humps are shown in Figs. 6, 7, 8, and 9.

Table 2: Pothole data

	Length (m)	Width (m)	Depth (m)
Pothole I	1	0.75	0.1
Pothole II	0.8	0.5	0.1
Pothole III-i	1.6	0.75	0.1
Pothole III-ii	1	0.5	0.1
Pothole IV	0.7	0.5	0.1

Table 3: LiDAR characteristics

	LiDAR data
Ray count	32
Measuring distance	200 m (1 to 200 m)
Rotation frequency	10 Hz
Field of view (horizontal)	360°
Field of view (vertical)	40°(+20° to -20°)

Figs. 6, 8, and 9 show the point cloud data acquired by the LiDAR system. The yellow dots are point data that were acquired by the LiDAR system and the white dots are point data that were identified as maintained surfaces. The blue dots are the point data that were identified as potholes and the red dots are the point data that were identified as humps.

Fig. 6b shows the road environment when Fig. 6a was acquired. Fig. 6b shows that the road in front of the vehicle is a maintained surface. The white dots in front of the vehicle in Fig. 6a indicate that the system is perceiving the maintained surfaces correctly.

Fig. 7b shows the road environment when Fig. 7a was acquired. Since the height of the gravel road is within LiDAR's error range, many parts of the road are perceived as maintained surface. The variance value of the variance of the height perceived as flatland is  $8.38 \times 10^{-4} \text{ m}^2$ . This variance value exceeds the set threshold. Therefore it is correctly identified.

Fig. 8b shows the road environment when Fig. 8a was acquired. The point data within the yellow rectangle in Fig. 8a are points that were identified as potholes. The yellow rectangle in Fig. 8b indicates that a pothole is in the same location as that shown in Fig. 8a. This result shows that the system is capable of perceiving potholes correctly.

Fig. 9b shows the road environment when Fig. 9a was acquired. The point data within the yellow rectangle in Fig. 9a are the points that were identified as a hump. The yellow rectangle shown in Fig. 9b indicating a hump is in the same location as that shown in Fig. 9a. This result shows that the system is capable of perceiving humps correctly.

These results show that the features of the road environment were identified correctly. The maximum distances from which potholes and humps can be identified by the LiDAR system when the vehicle is stationary are shown in Table 5.

Table 5 shows that smaller potholes are only identified when they are close to the vehicle, while larger potholes can be identified further away from the vehicle. Similarly, the longer



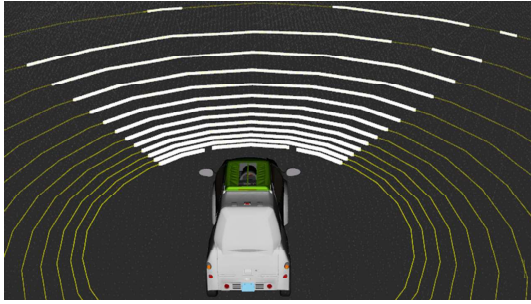


(a) LiDAR data for maintained surface



(b) Virtual environment used for maintained surface

Figure 6: Maintained surface identification results



(a) LiDAR data for gravel surface



(b) Virtual environment used for gravel surface

Figure 7: Maintained surface identification results

Table 4: Development environment

CPU	Intel®Core™i9-9900K CPU @ 3.6 GHz × 16
Memory	32 GB
GPU	NVIDIA GeForce RTX 2080
OS	Ubuntu 16.04 Xenial Xerus

a hump is, the more easily it can be identified by the autonomous vehicle.

Table 6 shows the distance from which the autonomous vehicle can identify a pothole or a hump when the autonomous vehicle is traveling at 10 km/h.

Table 6 shows that the distance from which obstacles can be identified varies depending on the size of the obstacle.

In addition, the road recognition times are shown in the table 7.

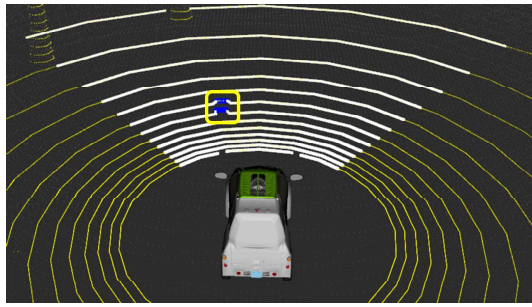
## 4 DISCUSSION

Autonomous driving systems must make appropriate decisions, even on undeveloped roads, such as private roads. In this study, experiments on the perception of the road environment by an autonomous vehicle system were conducted. The recognition distance was used as the main evaluation criterion. The maximum recognizable distances shown in Table 5 and 6 are the distances from the LiDAR system to each obstacle. The distance required to stop safely and comfortably is the sum of the safe stopping distance and the distance from

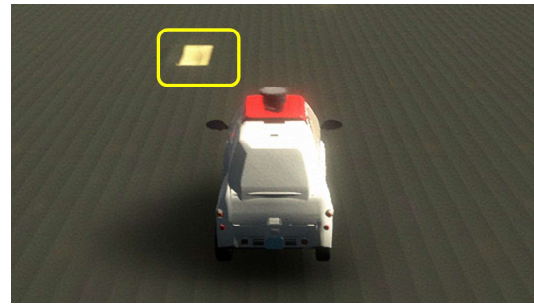
Table 5: Identification results (stationary)

Obstacle	Distinguishable distance (m)
Pothole I	8.26
Pothole II	6.04
Pothole III-i	8.24
Pothole III-ii	10.55
Pothole IV	11.5
Hump I	8.48
Hump II	9.51
Gravel surface	10

the LiDAR system to the front of the vehicle. The results showed that an obstacle has to be identified from a distance of more than 5.23 m. The obstacle recognition experiment performed while the vehicle was in motion showed that all of the objects, except for potholes II, could be identified at a safe stopping distance. A problem with the vehicle being unable to stop at a safe distance from an obstacle occurred when the pothole was located a short distance to the left or right of the autonomous vehicle. However, since a pothole that is not in the direct path of travel is not considered to be an obstacle, it does not need to be identified. Also when the car is traveling at 10 km/h from Table 7, it is possible to identify the road at a max distance of 0.18 m from the acquisition of LiDAR data. Therefore, even taking the recognition time into account, we believe that the car can stop safely at 10 km/h. Consequently,

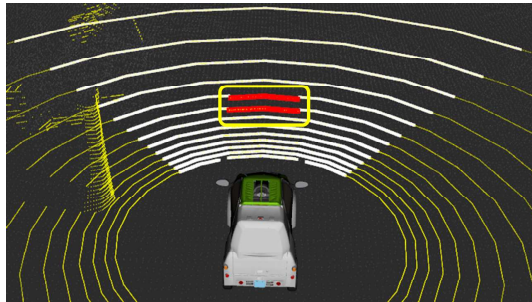


(a) LiDAR data showing a pothole

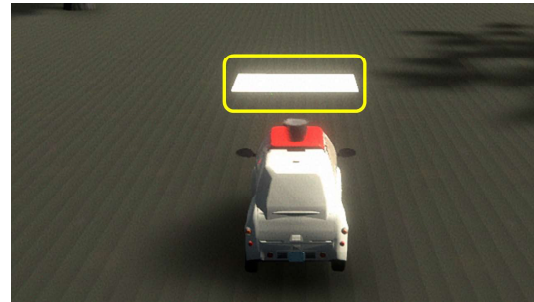


(b) Virtual environment showing a pothole

Figure 8: Pothole identification results



(a) LiDAR data for hump



(b) Virtual environment for hump

Figure 9: Hump identification results

Table 6: Identification results (while driving)

Obstacle	Distinguishable distance (m)
Pothole I	7.21
Pothole II	5.06
Pothole III-i	7.54
Pothole III-ii	6.31
Pothole IV	6.78
Hump I	9.11
Hump II	9.35

Table 7: Identified time

Vehicle status	Ave time	Max time	Min time
Stationary	1.38 ms	2.23 ms	1.15 ms
While driving	3.56 ms	65.3 ms	1.14 ms

we consider that the proposed method can safely control the vehicle after recognizing obstacles in the road environment.

In addition, the speed employed in this study was 10 km/h. If the travel speed is increased to 15 km/h, then the comfortable stopping distance would be 7.2 m. Consequently, when the distance of the LiDAR system from the front of the vehicle is considered, it is necessary to identify obstacles from a distance of more than 8.21 m. In such a case, the minimum distance for obstacle recognition should therefore be at least

8.21 m for all potholes shown in Tables 5 and 6. Thus, in the future, it will necessary to identify potholes from further away.

One of the reasons for the inability of the autonomous vehicle modeled in this experiment to identify obstacles at greater distances is the angular resolution of the LiDAR system. As shown in Fig. 2, the spacing of the LiDAR laser channels that was used to identify the road environment in front of the autonomous vehicle was relatively wide. As a result, small obstacles such as potholes are harder to identify the further they are from the vehicle. One solution to this problem would be to direct more of the LiDAR laser channel downward. Another method is to acquire point data at regular intervals using a LiDAR laser, which can identify the ground, and to identify the road environment from the changes in the height of the point data.

The present identification method has a limitation in identifying the road surface shape. If the pothole depth is too shallow, it is within the LiDAR's error range and is perceived as a maintained road. In addition, if the pothole is too small, it cannot be identified as a pothole because the multiple lasers of LiDAR cannot be identified consecutively. One solution is to acquire point data at regular intervals using a LiDAR laser, which can identify the ground, and to identify the road environment from the changes in the height of the point data. However when the pothole is shallow and small, I believe that there will be no problem with vehicle control.

In this study, we conducted recognition experiments in a virtual environment. But there are many issues that need to

be addressed when conducting evaluations in a real-world environment. For example, LiDAR in the real environment generates noise. Therefore, it is necessary to use LiDAR in the real environment to determine if it is possible to identify the road surface environment even in the presence of noise. In the future, it is necessary to conduct experiments using LiDAR in a real environment.

## 5 CONCLUSION

In this study, a method for identifying obstacles in the road environment was investigated to enable safe vehicle control for a given road environment. We created potholes, humps, and gravel surfaces in a virtual environment, characterized features in the road environment, and measured the maximum distance from which these obstacles could be safely identified. The results showed that potholes, humps, and gravel surfaces measuring at least 0.8 m long and at least 0.5 m wide could be detected from a minimum distance of 7.2 m from the vehicle. This minimum distance was sufficient for allowing the vehicle to decelerate and to stop comfortably after the obstacle was identified. For slow-moving vehicles, the proposed method successfully identified obstacles in the road environment and permitted safe operation of the vehicle.

We also intend to assess methods for recognizing obstacles in the road environment using less expensive LiDAR sensors. Since our experiments were conducted in a virtual environment, we intend to conduct experiments using a real vehicle in the future. In addition, we intend to develop an information provision method that can be used to create avoidance routes, and then run experiments that implement these routes. Another challenge is to devise a method for recognizing obstacles other than potholes and humps (e.g., gutters and curbs).

## Acknowledgment

This work was supported by Future Automotive Technology Development Project V6 from Knowledge Hub Aichi.

## REFERENCES

- [1] Mercedes-Benz. "Mercedes-Benz Innovation: Autonomous Driving." <https://www.mercedes-benz.com/en/next/automation/> (accessed on October 22, 2019)
- [2] Tesla. "Autopilot | Tesla." <https://www.tesla.com/jp/autopilot> (accessed on October 22, 2019).
- [3] Ort T, Paull L and Rus D. "Autonomous vehicle navigation in rural environments without detailed prior maps." 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, (2018).
- [4] Tukada M. "Development and Systematization of Human-Friendly Remote Operation Element Technologies: A Bridge to the Realization of Fully Automated Driving." [http://www.astf-kha.jp/project/project1/files/astf\\_PV\\_06\\_1011ol.pdf](http://www.astf-kha.jp/project/project1/files/astf_PV_06_1011ol.pdf) (accessed on June 4, 2020).
- [5] Isoda S, Kubota H, Sakamoto K, Aoki H. "Future Prospects for Hump Installation Standards in Our Country: Referring to Overseas Hump Installation Standards." Civil Engineering Society 57th Annual Academic Lecture, (2002).
- [6] Japan Road Association. "Examples of Asphalt Pavement Damage." <https://www.road.or.jp/event/pdf/201708213.pdf> (accessed on June 5, 2020).
- [7] Douroweb. "Accidents and response practices related to holes and bumps." [http://www.douroweb.jp/318administrative\\_fault/administrative\\_fault\\_pothole.html](http://www.douroweb.jp/318administrative_fault/administrative_fault_pothole.html) (accessed on June 4, 2020).
- [8] Sucgang, Nathalie Joy, Manuel Ramos Jr, and Nicolette Ann Arriola. "Road surface obstacle detection using vision and LIDAR for autonomous vehicle." Proceedings of the international multicongress of engineers computer scientists (IMECS), Hongkong, China. 2017.
- [9] Unity. "Unity for all." <https://unity.com/> (accessed on October 18, 2019).
- [10] LG Electronics. "LGSVL Simulator |An Autonomous Vehicle Simulator." <https://www.lgsvlsimulator.com/> (accessed on October 15, 2020).
- [11] Parallel and Distributed System Lab. "Automated driving software: Autoware." <https://www.pdsl.jp/fot/autoware/> (accessed on October 16, 2018).
- [12] Urano K, et al. "Road Surface Condition Survey using a Laser Scanner Mounted on an Autonomous Driving Car." multimedia, Distributed Coordination and Mobile Symposium 2018 (2018).
- [13] Open Source Robotics Foundation. "ROS.org." <http://wiki.ros.org/ja> (accessed on July 2, 2020).
- [14] ARGON CORPORATION. "Omnidirectional 3D-LiDAR sensor." <https://www.argocorp.com/cam/special/Velodyne/common/pdf/VLP-32.pdf> (accessed 2020/6/5)
- [15] Wang F, Sagawa K and Inooka H. "A study of the relationship between the longitudinal acceleration/deceleration of automobiles and ride comfort." Ergonomics 36.4 2000.
- [16] Ezawa M. "On the Beneficiation of Gravel and Crushed Stone." FLOTATION 1967.33 (1967).
- [17] National Institute of Advanced Industrial Science and Technology (AIST). "FY 2017 Research and Development and Demonstration Project for Social Implementation of Advanced Automated Driving Systems: A Report on the Implementation of a Terminal Transport System Using Autonomous Driving in Dedicated Spaces." 2018.



# Memory-Saving Software Update Method for In-Vehicle ECU

Kazuki Someya\*, Yoshiaki Terashima\*\*, Shunsuke Sugimoto\*\*\* Ryoza Kiyohara\*\*\*

\*Graduate School of Kanagawa Institute of Technology, Japan

\*\* Soka University

\*\*\*Kanagawa Institute of Technology, Japan

**Abstract** - Vehicles with autonomous driving, V2V, and V2R functions are being shipped to the market. These vehicles have many electronic control units and the size of software can be significant. Moreover, security risks are associated with being connected to the network. Thus, software update technologies by over-the-air technology are necessary. However, the bandwidth of most popular in-vehicle networks using controller area network technology is too small. This has time implications for updating software. Differential compression technologies have been developed for software updating. However, these technologies require large random-access memory (RAM) sizes, and there are some situations wherein the differential compression technologies cannot be used. In this study, we propose a new method applicable to small RAM-sized systems that offers improved general-purpose compression together with reuse of the original dictionary.

**Keywords:** software update, ECU, NOR type flash memory compression, binary difference.

## 1 INTRODUCTION

Many car manufacturers are developing autonomous vehicles. Autonomous vehicles have numerous electronic control units (ECUs) for sensors and controllers for equipment. The size and complexity of software are increasing over time, and, hence, releasing completely error-free software becomes increasingly difficult to achieve.

Moreover, the number of connected cars is increasing, and many ECUs attract attacks by hackers; therefore, regular software updates is essential [1,2].

Over-the-air (OTA) updating technology is appropriate for updating ECU software. In many cases, binary difference technologies are used for OTA updating, as this reduces the number of software that needs to be transmitted to correct errors in ECU software [3].

Figure 1 shows the typical software updating system for ECUs developed by many original equipment manufacturers (OEMs). The new versions of software are developed and maintained by OEMs and tested and authorized by car

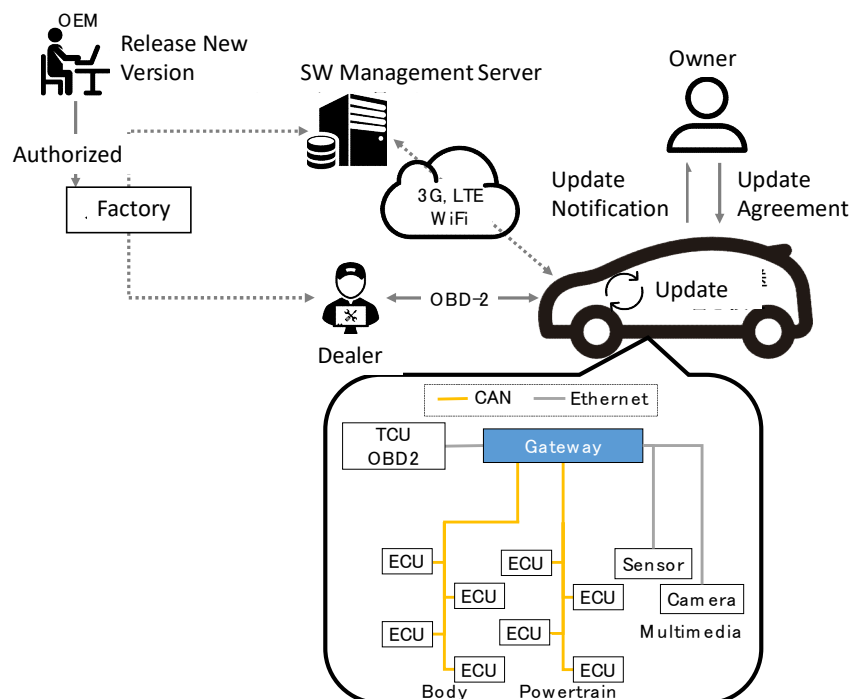


Figure 1: Software update for in-vehicle ECU.



companies. New versions of software are delivered by OTA technology through cellular networks or via direct connection to a second-generation on-board diagnostic (OBD II) port. The updated software is sent to individual ECUs through the in-vehicle network.

Typical in-vehicle networks are controller area networks (CANs) [4]. However, the network bandwidth is very narrow, at about 500 Kbps. There are faster in-vehicle networks [5], but they are quite expensive. The typical in-vehicle network, therefore, is slow, and most software updating time is consumed delivering the updates through the CAN [3]. Updating time depends totally on the update file size.

The driver cannot drive the car while the ECU software is being updated through the CAN. Therefore, the updating time must be as short as possible. For this reason, binary difference technology is critical [3,6].

When software changes are made, in many cases some code is deleted and added, as shown in Figure 2. There are also many cases in which some code is changed. However, this case can be processed by deleting the old code and adding the new code. As a result, some codes are moved. Therefore, binary difference updates can be represented by two commands.

However, even if the changed codes are small, there could be many differences in the code. Because there are many references such as a jump or accessing the memory, sometimes the references might be changed, as shown in Figure 3. This figure also shows the no-changed code being changed by the linker according to the references that are changed.

There are many studies on binary difference technology as discussed in Section 2.

There are many different types of ECU, some of which do not have sufficient random access memory (RAM) to use binary difference technology. Some ECUs have NOR-type flash memory that has a small amount of RAM. If the size of the RAM is smaller than the size of the erase block of flash memory, binary difference technology cannot be used. When part of the new software is written to the flash memory, the rest of the program in the same erase block cannot be referenced.

In many cases, fixing errors only requires small changes.

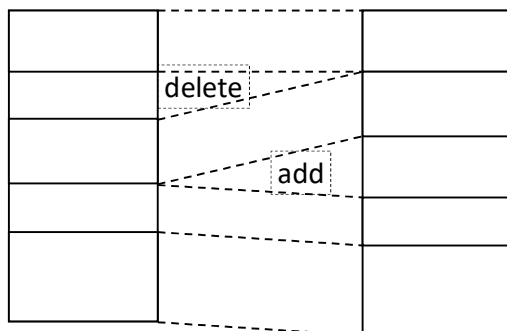


Figure 2: Binary Difference

However, there may be many differences in the binary code depending on the references. Therefore, effective compression technologies must be developed.

Many compression algorithms search for repeating sequences of symbols, as described in Section 2. The references are different in the different sequences. However, there are many of the same sequences that do not include the reference.

In this study, we focus on these cases and propose a new compression technique with a dictionary for extracting, and then we present an evaluation of our compression algorithm method.

## 2 RELATED STUDIES

There are many software update services available such as Windows Update and the software updating services used in mobile phones. However, these services can take a long time to execute, and this problem has been the focus of a large number of research. The problem consists of two main issues: time to deliver the new software version to the target device and time to rewrite the software on the target device.

Software updating for ECUs, satellites, etc. are processes that connect to the service through a slow network. For these cases, the delivery time is the largest problem. Software updating for cellular phones, personal computers, etc. are processes that often require a long time to rewrite the new version on the target device.

A large number of research focused on this area, including our previous studies [7-9]. Software structure is critical to the binary difference and rewriting time [7]. That study reported that the module structure is related to both the size of the binary difference and rewrite time. Some compression technology is useful for NAND flash memories because the decompression time is shorter than the time needed to read the non-compressed data from NAND flash memories. However, it is difficult to make small binary differences on such devices. Hence, we proposed a new binary difference technique for this type of device [8,9].

There are also several studies on data compression. Binary difference technology is the base strategy [3,10]. Many studies are based on binary difference technologies that require the RAM to be larger than the unit size of the

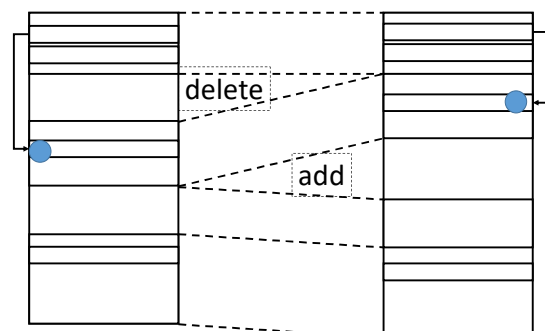


Figure 3: Many references are changed



erase block on flash memories. There are some studies that address reducing the RAM [11,12]. However, these techniques also require the RAM to be larger than the unit size of the erase block in flash memories.

Our previous study [13] assumed the RAM is small and cannot be used with binary difference technology. This study compares the general compression method from the time of extraction. There are many general-purpose compression algorithms [14]. From this study [13], the LZ77 algorithm [15] was found to be suitable for updating an ECU that cannot use binary difference technology.

There are two other aspects that should be considered. One is the integrity of the software. There are many integrity technologies suitable for this type of software, such as CRC check, and hash methods. The other aspect for consideration is security. There are many reports of attacks on vulnerable CANs. CAN communications are based on a message broadcast process. Therefore, if the attacker connects to the CAN through an OBD II port, they can easily sniff the messages. There are many studies related to this [16-20]. However, there are no studies about security combined with updating time. Therefore, in this study, we focus on the updating time.

### 3 SOFTWARE UPDATE FOR IN-VEHICLE ECU

#### 3.1 Software Update

In the factory, manufacturers update software to fix errors or address security problems. as shown in Figure 1. They know the following items for the target ECUs,

- CPU type and size of RAM for each CPU.
- Size of NOR flash memory and size of the erase block for each ECU

In the factory, they generate the data for updating the software from the old version to the new version. After testing the software, they release it. Vehicles download the new software by OTA with updates of ECU information.

If the ECU has sufficient RAM to be able to update using binary difference technology, the binary difference version is applied to the old version. Otherwise, the compressed form is applied for the update. We assume this type of software update.

#### 3.2 Software Update for ECU

Figure 4 shows how to update the ECU software. First, the new code in binary difference form is downloaded to RAM through the CAN. Second, the old version in an erase block is read from flash memory and stored in RAM. Third, the new code is applied to the old version and the new binary images are generated in the RAM. Then, the erase block on flash memory is erased. Finally, the new images are written into the erase block from RAM. In this process, the size of the RAM must be greater than the size of the erase block.

Therefore, if there is not enough RAM, it is not possible to use the binary difference technology to update the ECU. If there are spare erase blocks, they can be used by erasing and

writing the erase blocks twice. It might be care the writing time.

Figure 5 shows the flow for updating the ECU with general-purpose compression technology. The code for the update is downloaded and stored in the RAM through the CAN. After extracting, the data are written on the erase block that is cleaned. In this case, a large amount of code can be updated by repeating this process for different parts of the whole.

## 4 PROPOSED METHOD

### 4.1 Data Compression

There are many studies on data compression. Most studies are based on Huffman encoding [21], LZ77 [22], and LZ88 [23] and used a dictionary containing previously used strings. In our previous studies [13], many algorithms were compared based on different aspects, including rewired memory, time for extracting, and compression ratios. Our previous study showed that algorithms based on the LZ77 algorithm were suitable for updating ECUs. The compression time is not important because compression is processed in the factory.

In this study, we have improved the Zstandard algorithm

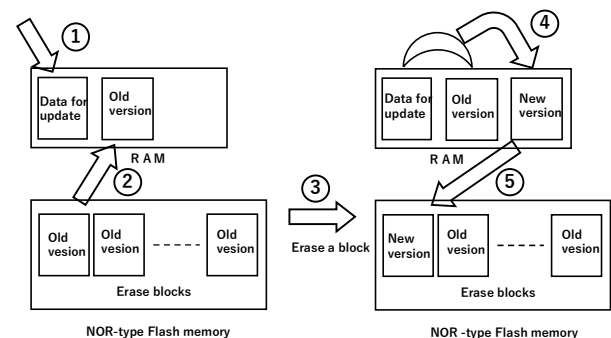


Figure 4: ECU software update by binary difference technology

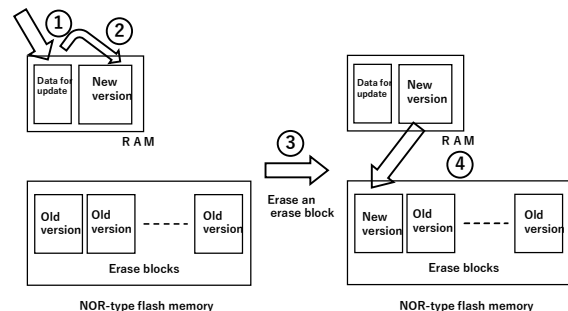


Figure 5: ECU software update by compressed data

[24] based on LZ77 for updating the ECU.

## 4.2 Reused Dictionary for Compression

LZ77 replaces the string position and length with details of previous occurrences of the same string. If there are no similar strings, it is not replaced. The area in which strings are searched is referred to as a window, which slides incrementally. The dictionary is made up from referenced strings.

Figure 6 shows the proposed method, which is an improved version of the original LZ77. Before shipping, the old version of the code is written on the flash memory. The old version is compressed, and the dictionary for extraction is also written on the flash memory and shipped with the old version.

When a software update is required, the new version of the software is compressed using the old (extraction) dictionary, which was saved with the old version on the shipped ECU. Then, after the compressed data are delivered, as shown in Figure 6, the old version is compressed. Then, the second string “ABC” is replaced with the position and length of the first string “ABC” in the dictionary. This position is not the string that appeared in the old version.

If software update is required, the new version is compressed with the old dictionary used for extraction. Certainly, the format of extraction and compression should be different. The dictionary for compression should have an index, etc. However, in this study, the two different formatted dictionaries were treated as the same dictionary because only the format was different.

The compressed old version with the pre-set dictionary (the old dictionary for extraction) was on the ECU, and the new data were decompressed using the pre-set dictionary. In Figure 6, the string “ABC” is in the pre-set dictionary. Then, the symbols that refer to the string “ABC” in the new code are replaced by “ABC.” With this method, the compression ratio is less than the conventional method, as shown in the lower part of Figure 6. The new symbol “D” was added to the old version. There are no symbols “D” in the pre-set dictionary. However, the conventional compression method

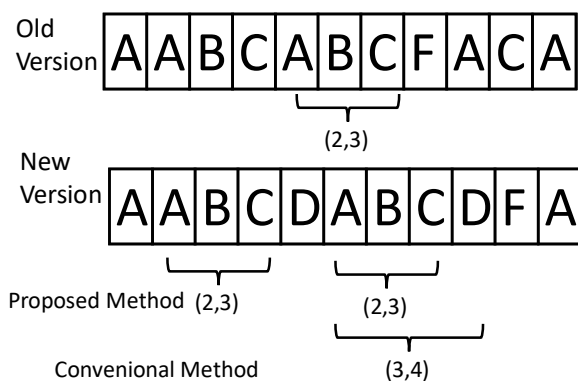


Figure 6: Dictionary for data compression

generates a new dictionary, which includes the symbol “D,” and can obtain a better compression ratio.

However, the download size is different. Only the compressed data without the dictionary have to be downloaded using our proposed method. The compressed data with the new dictionary must be downloaded by using the conventional method. This is the tradeoff between the two methods. Moreover, the proposed method requires additional flash memory. The size of flash memory is also a tradeoff between the two methods.

## 4.3 Flow of Software Update

Figure 7 shows the flow of the software update from initial shipping to generating the new software on the target ECU. We define the shipping version as the old version. Software update indicates that the old version is replaced by the new version. Before shipping, the old version is compressed by an algorithm based on LZ77 in the factory. Then, the dictionary part for extraction is extracted from the compressed data. The software code that is not compressed and the dictionary for extraction are written on the NOR-type flash memory in the shipped ECU. This dictionary is used for extracting the new data.

During the software update, the new version is compressed at the factory using the old dictionary. The compressed software without the dictionary is downloaded through OTA and the in-vehicle network. On the target ECU, the new software is extracted from the RAM using the (old) dictionary on the flash memory. The old code on the erase blocks is erased and the new code on RAM is written on to the blocks.

Figure 8 shows an example of the software update flow on the ECU. Our proposed method is suitable even if the size of the RAM is not large enough. In that case, the data for an erase block are divided into smaller portions, and the download, decompression, and update process are repeated for each portion. The flow is as follows:

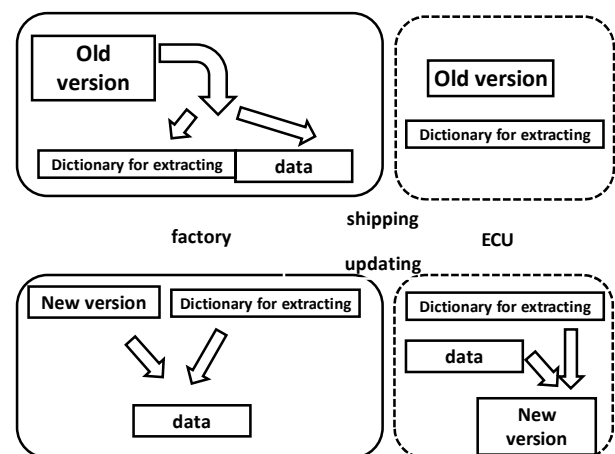


Figure 7: flow of software update

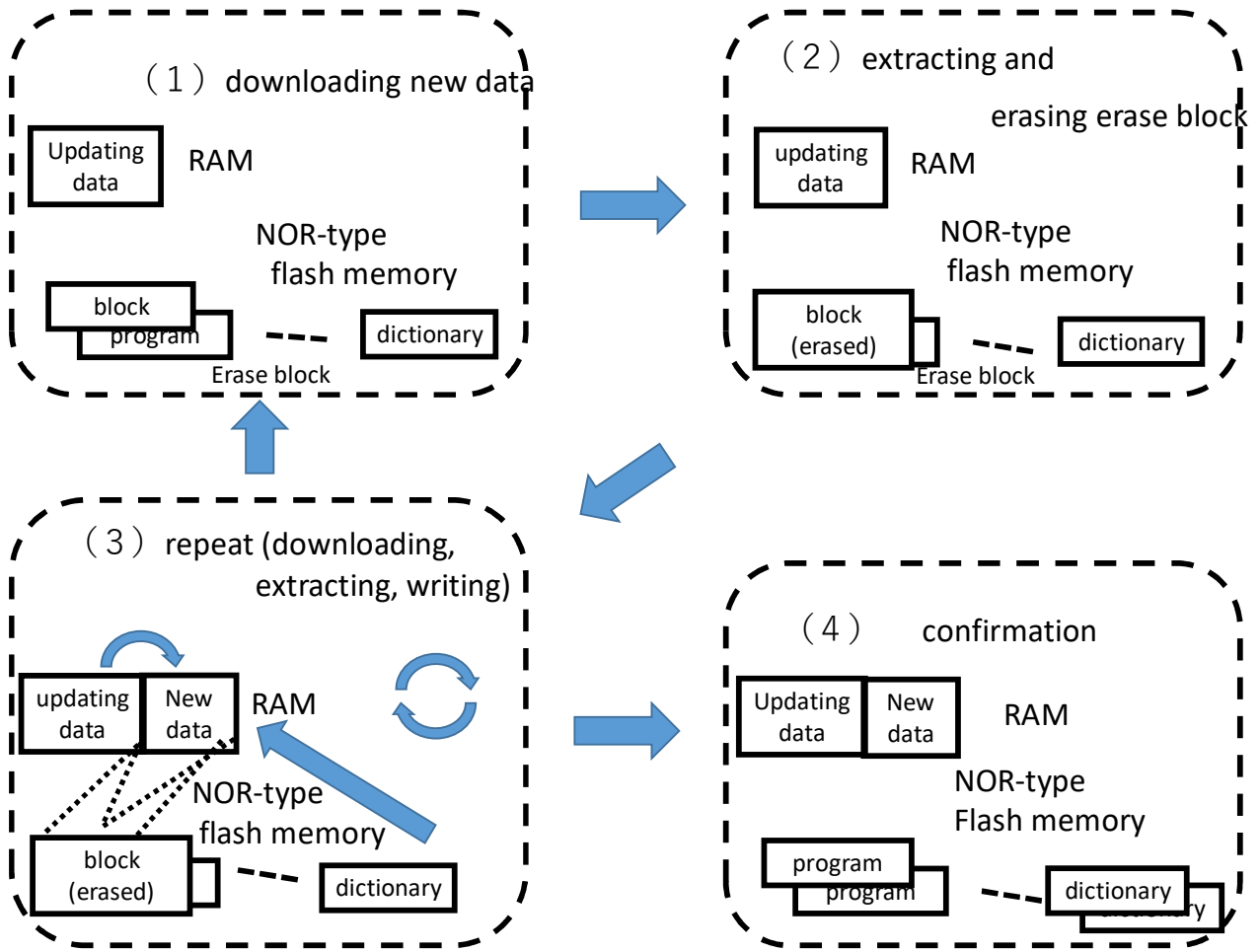


Figure 8: flow of software update

- (1) The first data is downloaded.
- (2) An erase block is erased.
- (3) The new data are extracted on to the RAM using the old dictionary, and the new code is written on the target block.
- (4) Steps (1)–(3) are repeated continuously until all the data are downloaded and extracted.
- (5) After the updating, the data should be confirmed by conventional methods (e.g., CRC or check sum).

## 5 EXPERIMENT AND EVALUATION

### 5.1 Experimental Conditions

We applied our proposed method to the Advanced Standard Profile kernel in Toppers [24] software, which is a well-known operating system for embedded CPUs. We prepared four versions that were numbered from 1.9.0 to

1.9.3. The size of the binary code for each version is shown in Table 1.

We built the binary code for the ECU that used the advanced RISC machine (ARM) processor. We assumed that the initial version is 1.9.0, which would have been the shipped version. Zstandard [25] was applied as the conventional compression algorithm.

We measured the compression ratio and compression ratio with multiple update for the conventional method and our

Table 1: data for evaluation (Toppers APS kernel)

Version	Data size(bytes)
1.9.0	28,742
1.9.1	28,742
1.9.2	28,746
1.9.3	28,746

proposed method

## 5.2 Evaluation

### 5.2.1. Compression Ratio

Table 2 shows the compression ratios after compression using the conventional method and the proposed method for each version. The size of the pre-set dictionary was 31,304 bytes. This dictionary included all strings and indices. Therefore, its size was larger than the original data size. However, the strings' pattern with no reference are necessary. These strings are needed for compression but not needed for extraction. Therefore, the compressed data excluded these types of strings.

The compression ratio is highly effective using the proposed method; however, it required additional flash memory space. This is the tradeoff. If the flash memory is too small, the size of the dictionary is limited. In these cases, short length strings or strings that are referenced only a few times are omitted to reduce the size of the dictionary. However, the size of the data is larger than the result with the original method.

### 5.2.2. Multiple Update

The software update for the ECU might be repeatedly applied. In the case of software for a PC or a mobile phone, new versions are released many times. Therefore, we should assume there will be multiple updates. If the binary difference technology is used, there are no problems because the difference between successive versions is generally very small. However, our proposed method uses the dictionary of the initial version. Therefore, the difference becomes larger in proportion to the version numbers. That is, updating data becomes more significant in proportion to version numbers.

Table 2 shows that the compression ratio increase is linear. However, update time might not be important in some situations. For example, a vehicle may be parked at the airport or with the dealer for maintenance. In this case, a long update time might be acceptable to the driver, and the new dictionary can be downloaded. Figure 9 shows the relation between version number and dictionary compression ratio.

This indicates that if the dictionary can be updated, the driver should update the dictionary.

Table 2: comparing compression ratio

Version No.	Conventional method(%)	Proposed method(%)
1.9.0	62.463	0.574
1.9.1	62.463	0.591
1.9.2	62.464	5.423
1.9.3	62.468	5.451

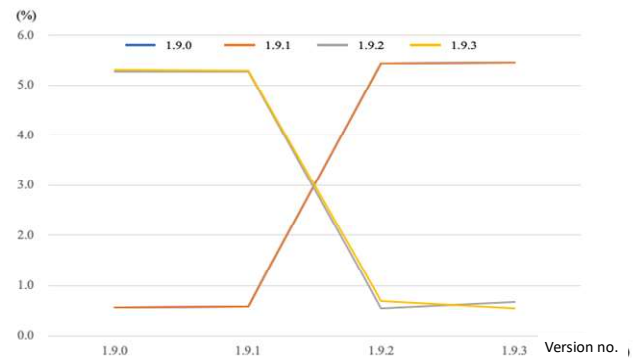


Figure 9 Dictionary size of multiple update

## 6 CONCLUSION

In this study, we have proposed a new compression method for updating ECU software. It is based on binary difference technology and conventional compression technologies. Many studies on software updating adopt binary difference technology. An ECU that does not have enough RAM cannot be used with binary difference technology. In that case, we propose a new method that improves the general-purpose compression technologies.

A suitable compression technology is LZ77, which our previous studies reported on. LZ77 makes a dictionary of references to previously appearing strings. We have proposed a new method that uses the old dictionary for extraction. We evaluated our method and achieved good results. However, we evaluated only with an ARM processor. We need to evaluate other platforms that will be part of our future studies.

## REFERENCES

- [1] S. Jafarnejad, L. Codeca, W. Bronzi, R. Frank, and T. Engel, "A car hacking experiment: When connectivity meets vulnerability," 2015 IEEE Globecom Workshops (GC Wkshps), pp.1–6, 2015.
- [2] A. D. Kumar, K. N. R. Chebrolu, R. Vinayakumar, and KP. Soman, "A brief survey on autonomous vehicle possible attacks, exploits and vulnerabilities," arXiv:1810.04144, 2018.
- [3] H. Teraoka, F. Nakahara, and K. Kurosawa, "Incremental update method for resource-constrained in-vehicle ECUs," IEEE 5th Global Conference on Consumer Electronics (GCCE) 2016.
- [4] R. Bosch, "CAN specification version 2.0", Robert Bosch GmbH, Postfach 300240, 1991.
- [5] S. Lorenz, "The flexray electrical physical layer evolution," SPECIAL EDITION HANSER automotive FLEXRAY, pp.14–16, 2010

- [6] D. Bogdan, R. Bogdan, and M. Popa, "Delta flashing of an ECU in the automotive industry," 2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI), pp.503–508, 2016.
- [7] R. Kiyohara, M. Kurihara, S. Mii, and S. Kino, "A delta representation scheme for updating between versions of mobile phone software," *Electronics and Communications in Japan*, Vol.90, No.7, pp.26–37, 2007.
- [8] R. Kiyohara, S. Mii, M. Matsumoto, M. Numao, and S. Kurihara, "A new method of fast compression of program code for OTA updates in consumer devices," *IEEE Transactions on Consumer Electronics*, Vol.55, Issue 2, pp.812–817, 2009.
- [9] R. Kiyohara, S. Mii, K. Tanaka, Y. Terashima, and H. Kambe, "Study on binary code synchronization in consumer devices," *IEEE Transactions on Consumer Electronics*, Vol.56, Issue 1, pp.254–260, 2010.
- [10] Y. Onuma, M. Nozawa, Y. Terashima, and R. Kiyohara, "Improved software updating for automotive ECUs: Code compression," *IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, pp.319–324, 2016.
- [11] R. Burns, L. Stockmeyer, and D. D. E. Long, "In-place reconstruction of version differences," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp.973–984, 2003.
- [12] T. Nakanishi, H. Shih, K. Hisazumi, and A. Fukuda, "A software update scheme by airwaves for automotive equipment," *International Conference on Informatics, Electronics and Vision*, pp.1–6, 2013.
- [13] Y. Onuma, Y. Terashima, Sumika, Nakamura, and R. Kiyohara, "Compression method for ECU software updates," *Tenth International Conference on Mobile Computing and Ubiquitous Network (ICMU)*, pp.1–6, 2017.
- [14] J. Uthayakumar, T. Vengattaraman, and P. Dhavachelvan, "A survey on data compression techniques: From the perspective of data quality coding schemes data type and applications," *J. King Saud Univ. Comput. Inf. Sci.*, 2018.
- [15] J. Ziv, and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp.337–343, 1977.
- [16] AUTOSAR, Requirements on Secure Onboard Communication, [https://www.autosar.org/fileadmin/user\\_upload/standards/classic/4-3/](https://www.autosar.org/fileadmin/user_upload/standards/classic/4-3/)
- [17] K. K. A. Gandhi, and C. Arumugam, "An Approach for Secure Software Update in Infotainment System," *Proceedings of the 10th Innovations in Software Engineering Conference (ISEC)*, pp.127–131, 2017.
- [18] Z. Qin, F. Li, Y-H. Wu, and C. Wang, "New ECU attestation and encryption mechanism for in-vehicle communication," *DEStech Transactions on Engineering and Technology Research*, (ssme-ist), 2016.
- [19] M. Steger, C. A. Boano, T. Niedermayr, M. Karner, J. Hillebrand, K. Roemer, and W. Rom, "An efficient and secure automotive, wireless software update framework," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 2181–2193, 2018.
- [20] T. Chowdhury, E. Lesiuta, K. Rikley, C-W. Lin, E. Kang, B. Kim, S. Shiraishi, M. Lawford, and A. Wassyng, "Safe and secure automotive over-the-air updates," Gallina B., Skavhaug A., Bitsch F. (eds) *Computer Safety, Reliability, and Security. SAFECOMP 2018. Lecture Notes in Computer Science*, vol 11093. Springer, Cham.
- [21] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp.1098–1101, 1952.
- [22] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp.337–343, 1977.
- [23] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Transactions on Information Theory*, vol. 24, no. 5, pp.530–536, 1978.
- [24] Toppers, <https://toppers.com> (accessed 2020/5)
- [25] Y. Collet, "Zstandard - Real-time data compression algorithm," <http://facebook.github.io/zstd/>, 2015.





# A Study of Patrol System by Automated Robotic Car

Yuto Nagai\*, Yuya Sawano\*\*, Ryozi Kiyohara\*\*

\*Graduate School of Kanagawa Institute of Technology, Japan

\*\*Kanagawa Institute of Technology, Japan

**Abstract** - Some robotic cars such as Toyota's Micro Palette have been developed. These small robotic cars are suitable for use as transport, guide, and patrol robots on campuses or in buildings. These small robots move autonomously using a location information system, obstacle detection system, and maps. We focus on a patrol system for finding suspicious but harmless people such as aged wanderer. We assume that the patrol robots have GPS for their location information system, laser imaging detection and ranging (LiDAR) for obstacle detection, an infrared sensor for human detection, and a Bluetooth device for authorized human detection. In this paper, we describe the results of our basic experiment using LiDAR and an infrared sensor.

**Keywords:** patrol, automated robotic car, LiDAR, Infrared sensor, detection of innocent suspicious person

## 1 INTRODUCTION

Numerous studies have recently been conducted on automated vehicles with many types of sensors, digital maps, and route guidance functions. Automated vehicles must avoid many types of accidents. Therefore, unmanned operated vehicles can operate in limited areas, and many studies on robotic vehicles that can be used within limited routes have been conducted [1–3]. However, new technologies for automated vehicles can reduce the many constraints placed on robotic vehicles. Therefore, some systems for robotic vehicles such as described in [1] have been developed. Such robotic vehicles are available for users within a limited area such as the inside of a campus or building.

The applications of such robot are as follows:

- (1) delivering services.
- (2) guidance services.
- (3) security services.

We are planning to introduce robotic vehicles on the campus of Kanagawa Institute of Technology to reduce costs. Through such an introduction, a delivery service can reduce its number of delivery and receiving personnel. Guidance services can use such vehicles to benefit their guests. Finally, security services can reduce their patrol costs.

In this study, we focus on security services such as patrols. The patrol service has the following objectives:

- (1) Detecting suspicious people who try to avoid a patrol (i.e., a deterrent).
- (2) Detecting and leading suspicious people who are not malicious.

- (3) Detecting suspicious objects.

However, it is difficult to distinguish authorized people from suspicious people who are non-malicious. In this paper, we propose a patrol system that distinguishes these two types of people. We then show the results of a basic experiment indicating that our proposed method can obtain good results.

## 2 PATROL

### 2.1 Patrol in KAIT Campus

In this section, we describe the aims of the patrol used on our campus. Figure 1 shows the KAIT campus. There are many buildings on campus. Buildings no. 2 and no. 5 are only used for lectures. Other buildings have lecture rooms, personal rooms for professors, and rooms for office workers.

During the day, the main purpose of a patrol is finding suspicious objects included illegal parking. Because there are many students, teachers, office workers, and neighborhood residents on campus during the daytime, they are not required to go through an admission and confirmation process to enter the campus. However, all rooms except lecture halls are locked, and only students and teachers with approval can enter them.

During the nighttime, only teachers and students with permission can stay on campus. Therefore, a patrol service is required. The purposes of such a patrol were mentioned in section 1. Currently, security patrols are provided outside the campus buildings several times at night. Moreover, they patrol inside the buildings for other purposes. Therefore, they cannot conduct a patrol if a problem occurs.

Thus, we studied a way to introduce patrol robots on our campus. The robot works as a delivery and guide robot during the daytime. At night, the robot works as a patrol robot. The main purpose of introducing such robots are to reduce costs. Another purpose is a safe patrol. When finding a person who has fallen, the patrol robot might have to put on a gas mask. If the robots have a gas sensor, they can send such information to the control center. Figure 2 shows the goal of our study.

The control center controls the parking, charging, and communicating space through WiFi. There are many buildings and many objects between buildings (e.g., benches and trash bins). We are planning to put 5G equipment between each building. In these areas, the possibility of finding a human is greater than in other areas. Therefore, robots should be able to be controlled by the center using 5G.

It is difficult to find suspicious people. However, robots may encounter non-malicious people. Because, they do not



Figure 1: Map of KAIT campus

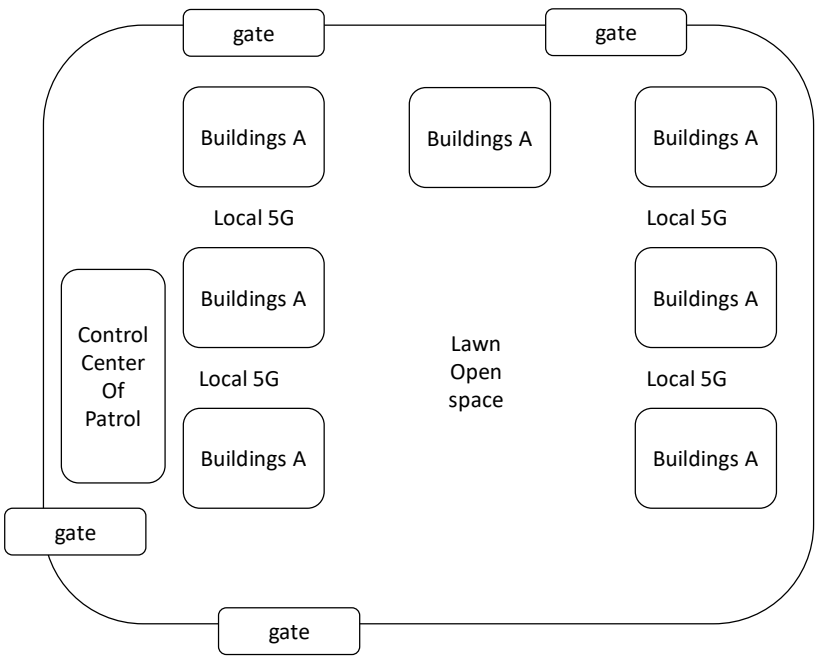


Figure 2: Image of goal of our study

think he is not illegal. Then, we hope that the robot can deter and find non-malicious people. Examples of non-malicious people include aged wanderers and children.

2.2 Current Patrol

Currently, many patrol systems consist of human patrollers and monitoring cameras. These cameras show the

surrounding area and information for security guards. Moreover, the camera is used for deterrence. However, security guards must decide what action they should take for every case when applying this system. This means that the guards are constrained by time. Moreover, it is not easy to analyze the still images from monitoring cameras at night or under poor weather conditions.

### 3 ROBOTS

We are planning to introduce robots for many different purposes. The robots are based on small automated vehicles, and have many sensors:

- GPS
- 2D and 3D LiDAR for object detection
- Infrared camera

Moreover, robots can communicate through a network such as LTE, WiFi, Bluetooth, and 5G.

We assume the robots are a type of autonomous vehicle. Therefore, they have a campus map. Moreover, they have 3D LiDAR (VLP-16) [4] and GPS. In addition, they can update the map dynamically using their own 3D LiDAR and can compare the previous information with the current information.

Moreover, we assume there are no steps outside of the buildings on our campus. The robot speed is almost the velocity of a walking human.

## 4 PROPOSED METHOD

### 4.1 Human Detection

Humans can be detected using 2D LiDAR. The robot can also have 3D LiDAR which has sufficient functions. However, the patrol might be an autonomous robot guided on the road by electricity or a magnet. Therefore, we propose using 2D LiDAR to reduce costs. Of course, if 3D LiDAR can be used, there will be no problems in applying our proposed method.

There are some methods used to detect humans through 2D LiDAR [5]. This technique depends on the fact that humans must move. If a human does not move, it is difficult to distinguish the human from an object. Therefore, we use an infrared camera as an additional sensor. This camera can easily sense the temperature of the target. Therefore, we combine the two types of data. It is also difficult to distinguish humans from other big animals (e.g., dogs and cats). However, on our campus, such cases are rare. In addition, the sensing length and width are different with both sensors. Therefore, we have to confirm this idea using our sensors.

### 4.2 Suspicious Object Detection

There have been many studies on suspicious object detection. Some techniques for object detection have applied improvised explosive devices without maps [6]. However, this technique requires a stereo camera. There are other techniques that detect using fast encoding of a point cloud [7].

In our proposed system, we adopt a conventional method using 2D LiDAR and maps. If the difference is larger than the threshold, the object is suspicious. However, we must implement and evaluate our system for defining the threshold.

### 4.3 Suspicious Human Detection

In our campus, some students and teachers walk around at night for relaxing or buying something. They are therefore not suspicious. However, we changed the rules for staying on campus at night. If a student or teacher wants to stay on campus at night, they have to apply to the office using an app on their smartphone. Then, if the request is approved, the app can receive the permission information.

The robots can communicate with a user's smartphone through Bluetooth and can confirm their permission. If they cannot confirm their permission, they are deemed suspicious. However, if the robot detects some people and not all of them are confirmed, it becomes difficult to identify the suspicious people.

However, at night, such cases are rare. Therefore, we should implement our system and test it in the field. If these cases are not as rare, the robot can go to each person and communicate through the NFC.

### 4.4 Proposed Experiment System

The robot is similar to a TOYOTA Micro-pallet [8]. Therefore, a 2D LiDAR (RPLiDAR A1M8) [9] can be installed on top of the robot (at a height of approximately 1 m). This LiDAR can scan 360 degrees. Table 1 shows the specifications of the 2D LiDAR. The maximum range of distance is 6 m. Although we think this distance is insufficient for the actual field, it is sufficient for our first step.

Table 1: Specification of 2D LiDAR (RP-LiDAR A1M8)

Item	Value
Size	98.5mm x70mm x60mm
Weight	170g
Range of distance measuring	0.15-6m
Range of horizontal angle	0-360
Resolution of distance	<0.5mm
Resolution of angle	≤1 degree
Sampling interval	0.5ms
Sampling rate	2000~2010Hz
Scanning rate	1~10Hz

Figure 3 shows our experiment system. The robot has a Raspberry Pi4 [10], which is independent from the autonomous vehicle functions. A Raspberry Pi4 has many communication functions including Bluetooth, WiFi, NFC, and LTE. Moreover, it can control many sensors such as LiDAR, an infrared camera, and GPS.

First, the 2D LiDAR detects an object or human and measures the angle and distance. Second, the infrared camera recognizes whether the target is human. Next, the Raspberry Pi4 searches for devices around the robot using Bluetooth. If a device answers, the application on the Raspberry Pi4 communicates with the application on the device and obtains the code information and confirms it with the server in the control center. If the code is authorized, there are no problems.

If there are more than two people, the Raspberry Pi4 communicates with same number of devices. If all devices are authorized, no problems occur. However, if a different number of devices is not authorized, the robot moves or calls another robot from a different location. Then, two or three robots will try to check the target area from a different location. As a result, the suspicious human can be identified, as shown in Figure 4. Moreover, if the human is difficult to distinguish, the robots approach the human and request the human to touch a smartphone with an NFC against the robot or request the human to input the code.

Figure 5 shows the data flow and operation of our system. Our system is a support system for a security guard. Therefore, it is not so severe to the accuracy. However, we must apply it in a test field for evaluation.

## 5 EXPERIMENT

We conducted an experiment on the functionality and accuracy of the 2D LiDAR as the first step of our system. Figures 6 and 7 show examples of a map of the area around the robots. Using this map, we can detect an object or human within approximately 6 m. This map is generated by a point cloud. The location is determined by comparing the dynamic map and static map with GPS. There have been many studies on this type of function [11–14]. Figure 6 shows a map with no people or walls. Figure 7 shows the results of the sensing of three people and walls. We can see the difference at the center of the map.

However, there are many cases of people and objects. Therefore, the suitable distance of the sensing differs for each case. As the second step, we have to experiment on our campus under numerous scenarios.

Moreover, we think suitable distance for measure is different for each situation. Thus, we have to try many kinds of LiDAR.

## 6 CONCLUSION

We proposed a patrol system and attempted to conduct an experiment as the initial step. We must conduct experiments on many aspects and using many types of sensors in the near future. Every year, at least one elderly person is found by a security guard on our campus. Every time, the guard brings them back to their home safely. However, we are afraid that an elderly person will have an accident on our campus. Therefore, security guards are expected to introduce our

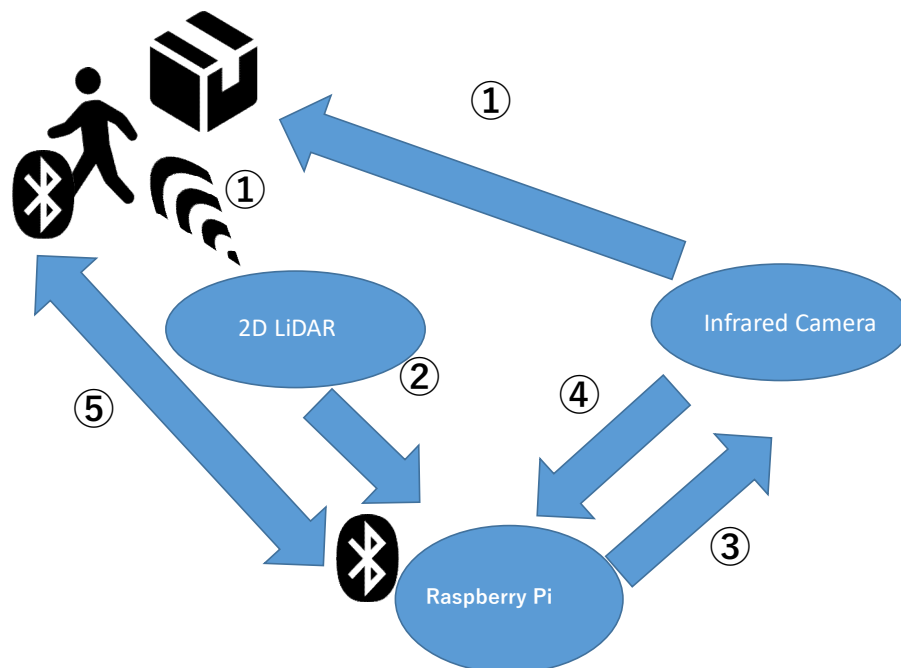


Figure 3: Proposed System

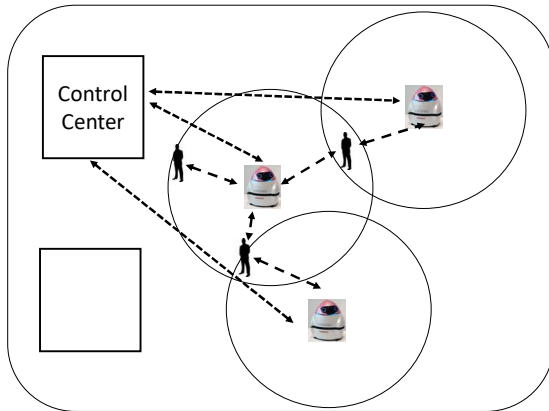


Figure 4: How to distinct the suspicious human and no-suspicious human

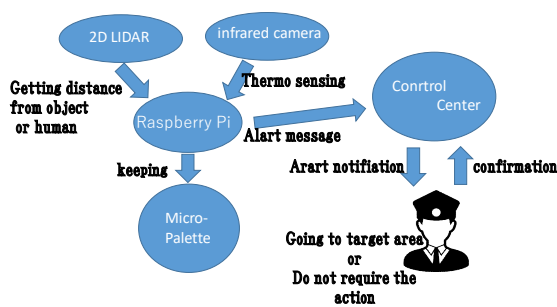


Figure 5: Data flow and operation in our Proposed System

robots into their patrols.

In this paper, we reported the results of the first stage of experiments. We plan to test the second stage of experiments in the near future.

## REFERENCES

- [1] SMP ROBOTICS: [https://smprobotics.com/security\\_robot/security-patrol-robot/](https://smprobotics.com/security_robot/security-patrol-robot/)
- [2] BIFUE USHIJIMA: <https://www.gov-online.go.jp/pdf/hlj/20181201/24-25.pdf>, Science and Technology, 2018
- [3] Security Robots on Patrol, <https://www.security-magazine.com/articles/89471-robots-on-patrol> Security Magazine
- [4] Velodyne: VLP-16, <https://velodynelidar.com/products/puck/>
- [5] Tapio Taipalus, Juhana Ahtiainen, "Human detection and tracking with knee-high mobile 2D LIDAR,"

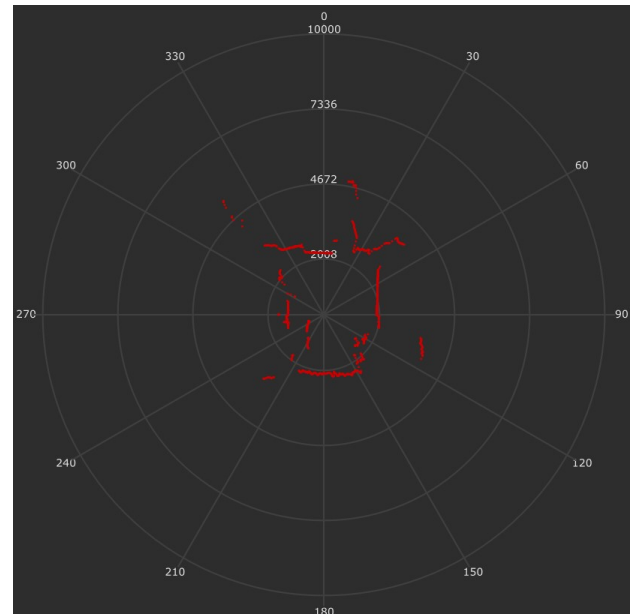


Figure 6: Example of map generated by RP-LiDAR

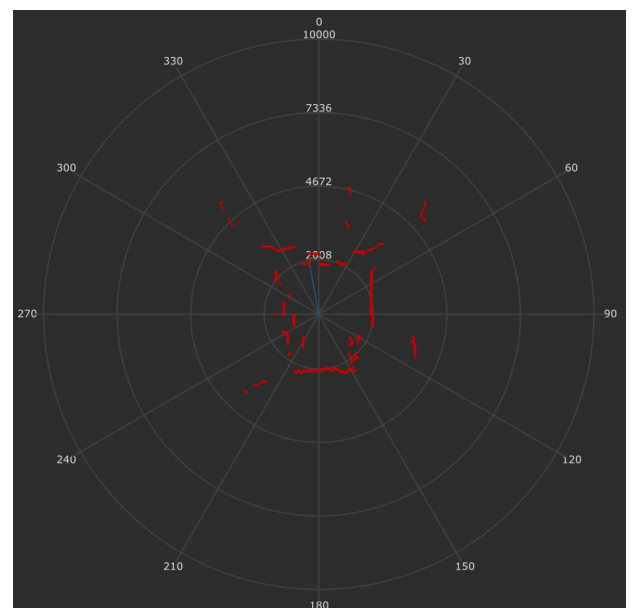


Figure 7: Example of map generated by RP-LiDAR with human.

- IEEE International Conference on Robotics and Biomimetics, pp. 1672-1677, 2011
- [6] Dennis W. J. M. , Gijs Dubbelman, Peter H. N. , "On improving IED object detection by exploiting scene geometry using stereo processing," Proceedings Volume 9407, Video Surveillance and Transportation Imaging Applications 2015.
- [7] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, Oscar Beijbom, "PointPillars: Fast Encoders for Object Detection From Point Clouds,"

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 12697-12705.
- [8] TOYOTA: <https://www.youtube.com/watch?v=lpLW34zch7o>
- [9] RPLiDAR, A1M8 <https://www.slamtec.com/en/Lidar/A1>
- [10] RASPBERRY PI FOUNDATION: <https://www.raspberrypi.org/>
- [11] Dongwook Kim, Taeyoung Chung, Kyongsu Yi, "Lane map building and localization for automated driving using 2D laser rangefinder," IEEE Intelligent Vehicles Symposium (IV), Seoul, pp. 680-685, doi: 10.1109/IVS.2015.7225763, 2015.
- [12] Daobin Wang, Huawei Liang, Tao Mei, Hui Zhu, Jing Fu, Xiang Tao, "Lidar Scan matching EKF-SLAM using the differential model of vehicle motion," IEEE Intelligent Vehicles Symposium (IV), Gold Coast, QLD, pp. 908-912, doi: 10.1109/IVS.2013.6629582, 2013.
- [13] Michael Bosse, Robert Zlot, "Keypoint design and evaluation for place recognition in 2D lidar maps," Robotics and Autonomous Systems, Vol. 57, Issue 12, pp. 1211-1224, 2009.
- [14] Z. J. Chong, B. Qin, T. Bandyopadhyay, M. H. Ang, E. Frazzoli, D. Rus, "Mapping with synthetic 2D LIDAR in 3D urban environment," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4715-4720, doi: 10.1109/IROS.2013.6697035, 2013



Session 6:  
Industrial and Agricarutural  
Applications  
( Chair: Hiroshi Inamura )



# High accuracy synchronized spindle/Servo control in the CNC Equipment

Kazuhiko Tsutsui<sup>\*†</sup>, Katsuhiko Kaji<sup>††</sup>, Katsuhiro Naito<sup>††</sup>, Naoya Chujo<sup>††</sup> and Tadanori Mizuno<sup>††</sup>

<sup>\*</sup>Mitsubishi Electric Corporation

<sup>†</sup>Graduate School, Aichi Institute of technology, Japan

<sup>††</sup> Aichi Institute of technology, Japan

<sup>\*</sup>[Tsutsui.Kazuhiko@ay.MitsubishiElectric.co.jp](mailto:Tsutsui.Kazuhiko@ay.MitsubishiElectric.co.jp),

## Abstract -

CNC machine tools are necessary to control the spindle and the servo axis in synchronization with each other, such as thread cutting and synchronous tapping. Conventionally, in case of thread cutting process and synchronous tapping process, it has been a general case that the control of the position and speed loop is carried out in the CNC side, and the configuration of the master-slave follow-up control is adopted in which the servo motor follows the spindle motor with slow response. However, as this method is based on the spindle's low response, there is a strict limitation to the high-speed and high-precision machining. Therefore, we first focused on developing the multi-core architecture to improve the characteristics of the spindle motor with low responsiveness. Next, we proposed a control method that improves the characteristics of the IM motor, whose characteristics tend to change in the temperature environment. Furthermore, we developed the compensation between amplifiers to reduce the synchronization error between the servo axis and the spindle caused by the difference in responsiveness, and verified its effect.

## 1 INTRODUCTION

CNC controller mainly generates the movement amount of each feed axis per unit time by analyzing sequentially the machining programs called G-code which describes the machining path of the tool, the feed speed of the tool and the number of revolutions of the tool. Further, a command value of the rotation speed of the spindle motor is generated according to the cutting conditions.

Here, the spindle motor is required to perform a milling process that requires an ultra-high speed of over 30,000 revolutions per minute. In addition, high power exceeding several tens of kilowatts to withstand heavy cutting is also required. Therefore, an induction motor (IM motor) that does not use a permanent magnet is used for the spindle motor, instead of a synchronous motor (SM motor) that uses a permanent magnet that is generally used for a servomotor.

Therefore, in the spindle motor control, it is difficult to realize high-response control at the same level as the servo. This means that it is difficult to completely synchronize the rotational position of the spindle and the position of the servo feed, such as thread cutting and synchronous tapping.

In the conventional these processes, CNC generates the feed axis position command as a master based on the rotational position of the spindle that response slowly and a relatively responsive servo axis follows the position of the slow spindle. In this method, it is necessary to limit that the processing speed is slow and acceleration is small.

In this paper, Section 3.1 of Chapter 3 proposes a method for

constructing a high-response, high-precision feedback loop that increases the responsiveness of the servo and the spindle itself and prevents the fluctuating in position and speed due to high-speed machining and cutting disturbances.

The most important point in constructing a high response feedback loop is minimizing the dead time and the processing cycle time in the loop.

Therefore, we have implemented position, speed, and current loop control in servo amplifiers and spindle amplifiers in order to minimize the effects of dead time caused by the communication interface outside feedback loop control.

In addition, in order to minimize the time delay and processing time in the servo and the spindle amplifier and to implement the function to improve the motor control performance in a specialized methods, a multi-core design was proposed, and the feasibility of high-speed and high-precision control of the spindle and the servo was verified.

Next, in Section 3.2, we proposed a control method that improves the characteristics of the IM motor, whose characteristics tend to change in the temperature environment, and stably achieves the performance of the spindle motor.

Finally, Section 3.3 proposes the method to realize compensation between amplifiers to compensate for the difference in response between the servo and the spindle that occurs even when the response of the servo and spindle is increased. As for the effect of this, Section 4 shows the verification results of this compensation. Section 4.1 describes in threading, and Sections 4.2 show the verification results in actual synchronous tapping.

## 2 RELATED TECHNOLOGIES

Chapter 2 shows the CNC system and the basic control structure of the servo and spindle, and then describes the problems of conventional spindle and servo synchronous control.

### 2.1 Basic Configuration of CNC Equipment

Multi-tasking machines such as the one shown in Figure 1 are increasing in recent years.

Each axes are driven by the servo motors connected to the ball screw in the machine. The tool used for cutting is attached to the spindle head and driven by the spindle motor. The CNC controller and the amplifiers which is used to control the speed and position in addition to power supply for the servo and the spindle motors are all installed in an electrical enclosure.

In the CNC system, since the path of the cutting tool directly affects the accuracy of the workpiece, it is important to follow the command with a small error against the influence of various load disturbances such as cutting disturbance and machine

friction.

It is also important to match the performance of synchronization and response between the X, Y, and Z axes. If the synchronization and response of each axis do not match, the tool path is not able to follow the command from CNC controller. It means that the processing(cutting) accuracy could not satisfy the required quality.

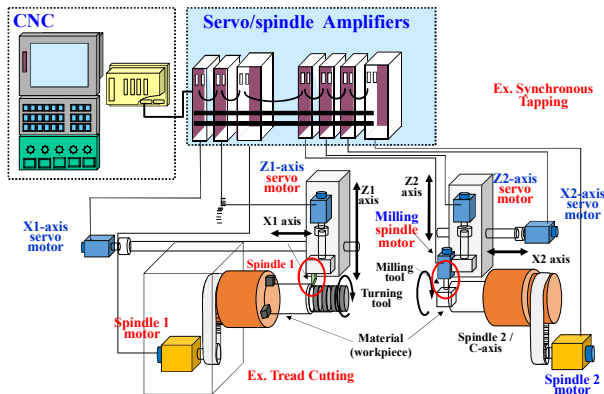


Figure 1: System configuration example of multi-tasking machine

## 2.2 Basic Servo/Spindle Control Architecture (Distributed Control)

Figure 2 shows a control block diagram at the time of thread cutting in which the rotational position of the main spindle and the position of the servo axes need to be synchronized.

Permanent magnets are used for servomotors that require high-response position control, while induction motors that do not use permanent magnets are generally used for spindle motors. For example, in a machining center or lathe that uses CNC, it is possible to rotate a rotating tool more stably higher speed than a case that requires position control of a spindle motor, or to withstand heavy cutting well. Since it is important to achieve the highest possible output, synchronous motors that are not suitable for high-speed rotation or high-output at high-speed rotation are rarely used.

On the other hand, machining required for machine tools is becoming more complicated, and machining that requires synchronizing the positions of the spindle motor and the servomotor, such as threading and synchronization taps, must be performed by one machine.

As described above, the response of the spindle motor is low which means that the response frequency is one digit lower than servo motors. Therefore, the method of centralized control as shown in Figure 2 that the servo axis using the position feedback of the spindle motor which is returned to the CNC side as a position command of the servo axis in such thread cutting is common in general.

As described above, the servo follows the movement of the spindle having low response, so that the synchronization error between the spindle and the servo can be reduced.

## 2.3 Issues with Spindle and Servo Synchronous Control

In this section, the problems in the conventional synchronous control between spindle and servo are shown by way of example of thread cutting control and synchronous tap machining.

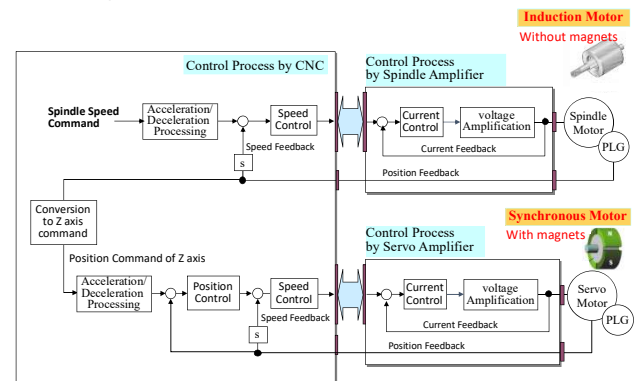


Figure 2: Synchronous control between Spindle and Servo (Thread Cutting)

### 2.3.1 Thread Cutting Control

It is possible to improve the synchronization performance between the spindle and the servo axis by constructing a control method as shown in Figure 2 and making the servo with relatively quick response follow the position of the spindle motor with low response. Since a spindle with low response is used as a reference, there is a problem that unless the response of the spindle increases, the tact time cannot be improved by increasing the speed of thread cutting.

In addition, the position feedback of the spindle motor is returned to the CNC side once, passed through the process of converting to the servo axis command, and then passed again to the servo amplifier via the network, which increases waste time. Therefore, there was a limit to suppresses synchronization errors eventually. Further, since there is a difference in the acceleration / deceleration time between the spindle and the servo, machining cannot be started until the spindle and the servo reach a constant speed in order to ensure synchronization accuracy. Alternatively, there is also a problem that the speed of the main spindle and the X-axis cannot be reduced the speed until the Z-axis finishes to raise as shown in Figure 3.

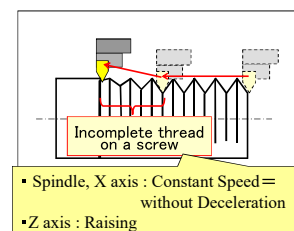


Figure 3: Incomplete thread on a screw

### 2.3.2 Synchronous Tapping

Also, in synchronous tapping, similar to thread cutting, since it is necessary to control the servo axes in synchronization with the rotational position of the main spindle, centralized control on the CNC side is adopted in some cases. On the other hand, compared to the thread cutting, since the drilling is performed after the preliminary hole is formed in advance, the cutting load is relatively small, and unlike the thread cutting which rotates the work itself, the inertia of the cutting tool (tapper) itself is small. In some cases, position commands to the spindle and the

servo are given independently of each other, and in many cases, the configuration is aimed at higher-speed machining.

However, if there is a slight deviation in synchronism such as the influence of the bite of a facet, there are problems such as inducing damage to the tapper and making it impossible to process a screw conforming to the standard.

### 3 PROPOSED SYSTEM

#### (High-Precision Synchronous Control System between Spindle and Servo)

This section describes the proposed method in this chapter to realize the synchronous control between the spindle and servo with high response and accuracy. The methods are largely composed of the following three methods, the basic structure of which is shown in Figure 4.

As mentioned above, in the conventional system, there have been many cases in which the controller controls the position and speed loop and switches the control method according to the machining process, especially in the threading and synchronous tapping in which synchronous control between the spindle and servo is required. In this paper, in order to pursue the high-speed command follow-up characteristics and the robustness against disturbance, we propose the distributed control system in which the position and speed loop are executed by the servo and main spindle amplifier.

Based on this decentralized control, we could achieve the high-precision synchronous control system between the spindle and the servo by using the following three methods.

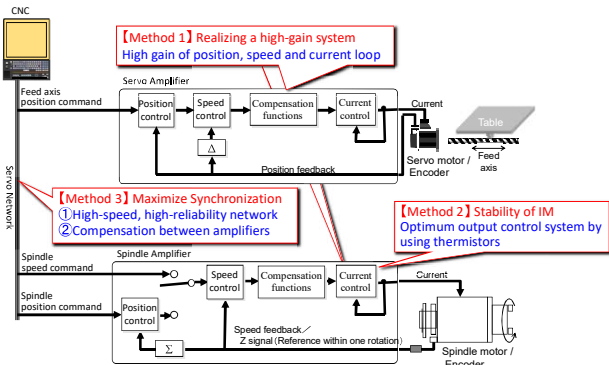


Figure 4: Basic configuration of high response synchronous control between spindle and servo

#### [Method 1] High gain of position, speed and current loop

Section 3.1 describes the details of Method 1 for realizing a high-gain system in which each independent spindle and servo follows commands from the controller at high speed and with high response and is not easily affected by cutting disturbances.

#### [Method 2] Power optimization of IM spindle motor using thermistors

Induction motors (IM motor) used in the spindle are greatly affected by temperature characteristics. Because it is generated by the motor's own coil and magnetic circuit, which is different from a synchronous motor in which the magnetic field for generating torque can be obtained from a stable permanent magnet.

Therefore, we have developed an optimum output control system using a thermistor, and its effect is described in Section 3.2.

#### [Method 3] High-speed, high-reliability network and compensation between amplifiers

Section 3.3 describes compensation between amplifiers to maximize synchronization performance by complementing the difference in responsiveness between the spindle and servo, and also describes high-speed, high-reliability networks to realize this.

### 3.1 PROPOSED METHOD 1

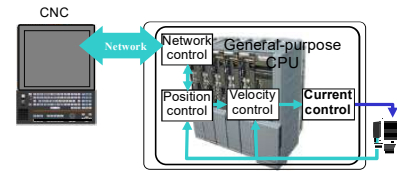
#### (High gain of position, speed and current loop)

Figure 5 shows a block diagram of the newly developed the current control core unit. Conventionally, the processing of the position, speed, and current loop in the spindle and servo amplifier has been sequentially performed by software processing by a general-purpose CPU.

On the other hand, in order to greatly improve the performance, we have developed a dedicated calculation unit which is called the current control core unit that is the innermost among the control loops and requires the highest gain (high response).

Thus, oversampling function is able to achieve the system that position/speed loop processing and the current loop processing could run in parallel, so that the PWM switching frequency can be stably increased, and high gain of the position/speed/current loop can be realized.

#### Conventional control: Software base control



#### Developed control: Current control core unit = High gain control system

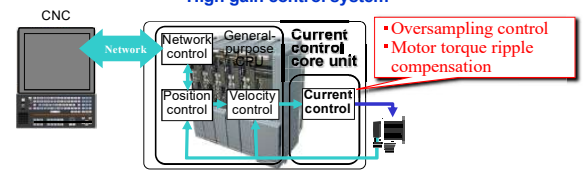


Figure 5: Current control core unit

### 3.2 PROPOSED METHOD 2

#### (Output Optimization of IM Main Spindle Motor Utilizing Thermistors)

The torque  $\tau$  of an induction motor (IM motor) generally used for a spindle motor is given by the following equation.

$$\tau = \frac{PM}{2L} i_q \psi_d$$

$$\psi_d = M i_d$$

$\tau$  : Torque of the induction motor

P : Number of poles

M : Mutual inductance between the stator and rotor

$L$  : Inductance

$\psi_d$  : interlinkage flux

$i_q$  : q-axis current ,  $i_d$  : d-axis current

Here, since the interlinkage magnetic flux is made of the product of the d-axis current flowing in the motor and the inductance instead of a permanent magnet like a servo motor (Synchronous motor), it is easy to be affected by temperature, and it is difficult to always obtain a stable output.

Therefore, as shown in Figure 6, we developed the system that a thermistor is built into the coil of the spindle motor, and it enables the spindle amplifier to constantly monitor the temperature. In this system, stable output characteristics are obtained regardless of temperature by controlling the d-axis current  $i_d$  according to temperature conditions.

### 3.3 PROPOSED METHOD 3 (Compensation between Amplifiers)

In Section 3.1, we describe how to improve command followability and rigidity against external disturbances by realizing high gain control of each of the servo and spindle.

However, in the spindle control, there are cases where the command follow-up characteristics are inferior to the servo because an induction motor (IM motor) which is difficult to achieve high responsiveness due to the influence of an electrical time constant and high inertia is used and the

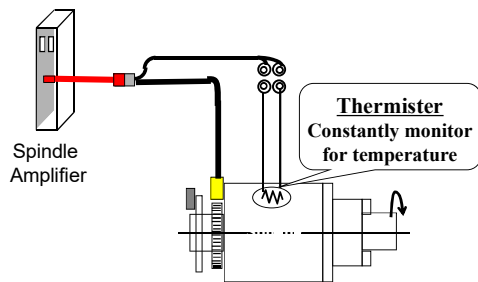


Figure 6: Constant monitor for temperature

acceleration/deceleration characteristics are not stable due to voltage saturation (torque saturation) caused by a power source environment or the like. It means that it is sometimes difficult to obtain high synchronization performance between the spindle and the servo axis.

Therefore, we have developed a high-speed synchronous tap function that utilizes the data communication protocol between amplifiers incorporated in a high-speed optical network and utilizes the amplifier compensation between the spindle and servo.

Figure 7 shows a block diagram of the high-speed synchronous tapping function. A position command synchronized with the spindle and the servo is sent via a network. On the other hand, the spindle amplifier and the servo amplifier perform position loop control to control the motor to follow the command.

If it is possible to ideally follow the command without any cutting disturbance or the like, synchronization accuracy can be guaranteed, but in general, it is difficult to improve the speed frequency response of the spindle motor control more than that of the servo control shaft.

Therefore, in the proposed method, the position deviation and the speed feedback in the spindle amplifier are passed to the

servo amplifier using the inter-amplifier data reception protocol provided in the high-speed optical network. On the servo amplifier side, unit conversion is performed using the spindle position sent by the CNC controller and the servo axis conversion coefficient  $K$  which means a ball screw pitch, and the spindle position deviation is added to the servo axis position command as a correction position, and the spindle speed feedback is added to the speed command as a correction value. The spindle speed feedback is further differentiated, converted by the inertia  $J$  of the servo and the torque constant of the motor, and then added to the current command as a correction current value. Since these correction position commands and current commands are data delayed by the network, they are values fed forward to make up for the dead time  $T_d$  caused by the network.

As a result, it is possible to achieve high precision and high speed in the servo-to-spindle synchronous control such as the synchronous tapping compare with the conventional control, and this can greatly contribute to the improvement of productivity.

On the other hand, we have also developed a high-speed optical communication servo network that dramatically improves the network performance between the CNC-servo and the spindle.

The points of this network are as follows.

- ① Communication baud rate improvement:  
5.6 MHz  $\rightarrow$  50 MHz (Approximately 10 times)
- ② Improvement of communication cycle:  
1.7 msec  $\rightarrow$  0.2 msec (about eight times)
- ③ Improvement of command resolution: 1  $\mu$ m  $\rightarrow$  1 nm
- ④ Protocol that enables data exchange between servos and spindle

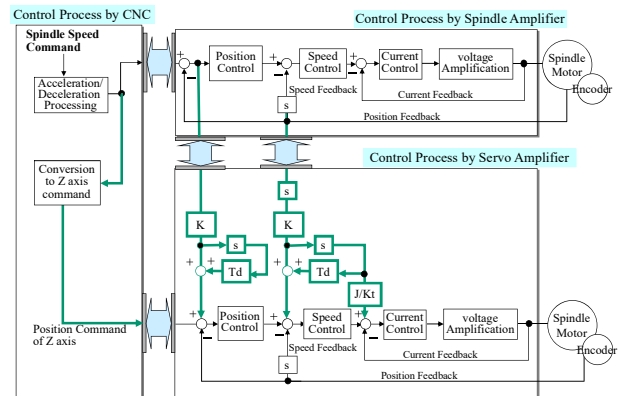


Figure 7: Diagram of High-speed synchronous tapping

## 4 EVALUATION

### 4.1 Thread Machining Improvements: (Simulation Validation)

Figure 8 shows the simulation results of threading by synchronous control of the spindle and servo. This time, the conventional system in which the servo follows the rotational position feedback of the spindle has been changed to a system in which the spindle and the servo follow the position command. Synchronization accuracy is improved by performing compensation between the spindle and servo in conjunction with high gain control.



Besides, conventionally, in order to ensure synchronization accuracy, machining cannot be started until the spindle and servo reach a constant speed, but in the proposed method, the synchronization accuracy can be improved even when the spindle speed is changed.

As shown in Figure 9, the spindle can be decelerated (Override) when the Z-axis tool is pulled up, and the length of the incomplete threaded portion can be shortened.

#### 4.2 Verification of Synchronous Control Performance between Spindle and Servo (High-speed and High-precision Synchronous Tapping: Verification of Real Machine Machining)

In the normal control, the synchronization error increases as the acceleration/deceleration time constant of the spindle decreases, and especially when the time constant is set so as to be small as to enter the torque saturation region, the increase of the synchronization error becomes remarkable. In other words, in actual use, it is necessary to set the time constant with enough margin so that torque saturation does not occur in consideration of changes in the power supply environment. On the other hand, when the compensation between amplifiers is enabled, the synchronization error is suppressed even if it is applied to the torque saturation region. It means that it is not necessary to set the time constant with enough margin as in the past, and it can greatly contribute to shortening the machining time.

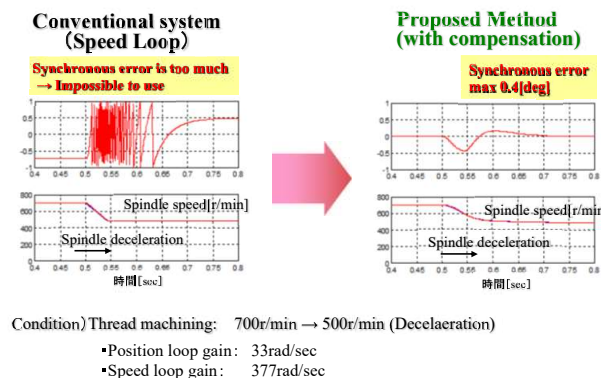


Figure 8: Synchronous accuracy in case of thread machining (Overriding speed condition)

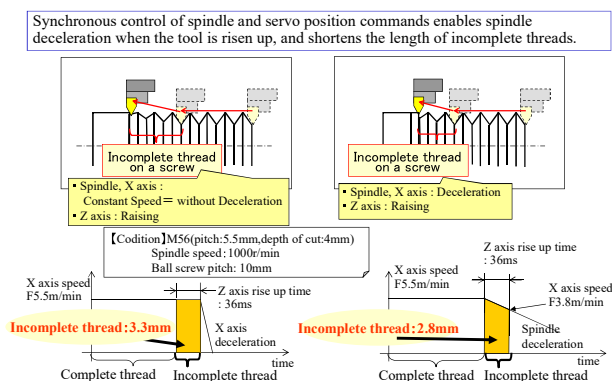


Figure 9: Reduction of incomplete thread

Figure 10 shows the results of continuous machining of

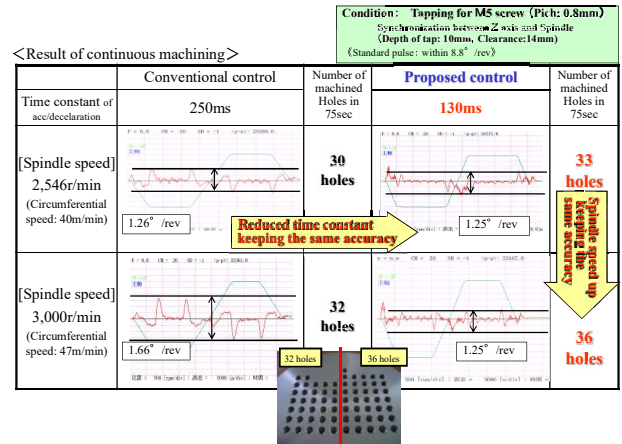


Figure 10: Result of continuous tapping by proposed control

synchronous tapping of M5 screws on an actual machining center. Even if the spindle speed and acceleration/deceleration time constant are shortened while securing the screw accuracy under the conventionally set machining conditions without the compensation between amplifiers, the accuracy can be secured by performing the compensation between amplifiers, resulting in productivity improvements of 30 to 36 units and 20% within the same time (75 seconds).

## 5 CONCLUSION

In recent years, in order to improve productivity, the demand for combined machining corresponding to various machining operations by one machine tool has been an increase. Under these circumstances, there has been an increase in the number of cases that the spindle and the servo are machined in synchronization. This paper introduces the methods implemented to realize high speed and high precision in the synchronous machining of the spindle and servo.

At first, we proposed a method for constructing a high-response, high-precision feedback loop that increases the responsiveness of the servo and the spindle itself and prevents the fluctuating in position and speed due to high-speed machining and cutting disturbances.

Specifically, in order to minimize the time delay and processing time in the servo and the spindle amplifier and to implement the function to improve the motor control performance in a specialized method, a multi-core design was proposed, and the feasibility of high-speed and high-precision control of the spindle and the servo was verified by showing bellows. In the case of the spindle motor control, the effect of reducing heat generation of the spindle motor for suppressing deterioration in accuracy due to thermal expansion of the machine is shown. In the case of the servo control, the effect of improving the accuracy of the feed shaft is shown.

Next, we proposed a control method that improves the characteristics of the IM motor, whose characteristics tend to change in the temperature environment, and stably achieves the performance of the spindle motor, and verified its effectiveness for the fluctuation suppression of acceleration and deceleration time in a real machine.

Finally, we proposed the method to realize compensation

using the high-speed and high-reliability network between amplifiers to compensate for the difference in response between the servo and the spindle that occurs even when the response of the servo and spindle is increased. As for the effect of this, we showed the simulation verification results in threading, and showed the verification results in actual synchronous tapping.

As a result, in comparison with the conventional control, for example, under the conditions of the synchronous tap processing shown in this document, productivity improvement of 20% was achieved. Besides, it is able to make a robust system to power supply environment.

This also served as a guideline for making the most of the future evolution of semiconductor processes for spindle and servo control, and at the same time showed the importance of the network between the CNC controller and spindle and servo.

In the future, we plan to continue to optimize the architecture and compensation algorithms to maximize the synchronization performance of the spindle and servo, while also incorporating the evolution of semiconductors and network technologies timely.

## REFERENCES

- [1] Toshiki Tanaka, Kazuhiko Tsutsui : Driving system MDS-D/DH series for M700, Mitsubishi Electric Advance, Vol.116, pp. 23-25(2006.12).
- [2] Hidehiko Sugimoto, Masato Koyama, Shinzo Tamai : The theory of AC servo system and real design, General electronic publisher, p.31-71, 1997.
- [3] Shinichi Furutani, Akira Satake, Kazuhiko Tsutsui : Proposal of current control for high speed AC motor control, Journal of the Institute of Electrical Engineers, D, Vol.128, No.12, p.1361-1402.
- [4] Shinichi Furutani, Akira Satake : Current control method of AC motor for low carrier frequency drive, Journal of Lectures at the National Congress of the Institute of Electrical Engineers, p.224-228, 2006.3.
- [5] R. Toutant, S. Balakrishnan, S. Onyshko, and N. Popplewell : Feedrate Compensation for Constant Cutting Force Turning, IEEE Control Systems Magazine, Vol.13 , No. 6 , 1993.12.
- [6] Dong-II Kim, Chung Hyuk Yim : All digital high performance controller for spindle motor in CNC machine tool, IEEE International Electric Machines and Drives Conference Record, pp. MC2-2.1-MC2-2.3, 1997.
- [7] Yuji Nakamura, Shigeru Futami : Application of predictive control for FA mechatronics equipment, Journal of the Society of Instrument and Control Engineers, Vol.39, No.5, 2005.5.

# Development of a Yield Prediction Model Generation Process for Fruits and Vegetables in a Plant Factory

Yuki Todate\*, Michiko Oba\*\* and Mitsuru Takamori\*\*\*

\*Graduate School of Systems Information Science, Future University Hakodate, Japan

\*\* School of Systems Information Science, Future University Hakodate, Japan

\*\*\* Apure Inc, Japan

{g2119026, michiko}@fun.ac.jp

*Abstract* – In food plant factories, data are collected using production management systems and external and internal environmental sensors. However, predicting yields for fruit and vegetable crops is difficult because they are often cultivated outside and, therefore, are frequently affected by unstable external environmental conditions. Thus, analysis of yield-related factors and prediction have been carried out at various plant facilities. Explaining the unifying factors is difficult because each facility has different equipment and environmental control methods. Therefore, the purpose of this study is to identify the factors related to fruit and vegetable yields according to facility-specific conditions and to develop a model generation process to predict the yield that is applicable to various facilities. The model generation process involves selection of data, feature design and preprocessing, selection of model structure, and optimization and evaluation of the model. To identify the model structure, Multiple Linear Regression Analysis, Generalized Additive Models, and Random Forests are used.

*Keywords:* Yield Prediction, Plant Factory, Fruits and Vegetables, Statistical Modeling, Machine Learning

## 1 INTRODUCTION

The number of food plant facilities worldwide continues to increase and is estimated to continue to increase [1][2]. According to a study by the Yano Research Institute, the global market size of facility-grown crops is expected to be 16.5 billion yen in 2020 [3]. The number of plant facilities has increased due to the frequent occurrence of unseasonable and abnormal weather conditions and the growing safety and security consciousness of consumers.

A plant factory is a cultivation facility that enables the planned production of vegetables and other plants through advanced environmental control and growth prediction [4]. In such factories, to facilitate growth analysis and yield prediction, significant amounts of data are collected using production management systems and external and internal environmental sensors. Yield prediction is important to match market demand and prevent overproduction. However, yield prediction for fruits and vegetables, such as eggplant and tomato, has been difficult [2][5].

Two main factors make it difficult to predict the yield of fruits and vegetables. First, such crops are often affected by unstable external environmental conditions. In general, plant factories are classified as artificial light plant factories and

solar powered plant factories, which rely on sunlight [6]. Because fruits and vegetables require strong light for growth, they are commonly grown in solar plant factories. Therefore, growth is affected by weather conditions, such as the rainy season and winter. Since plant windows are opened and closed to prevent the occurrence of mold in the facility and the air is taken in from outside, it is also affected by the external environmental factors other than available sunlight, such as temperature and humidity. Second, yield prediction is related to individual growth characteristics. For example, compared to leafy vegetables, fruits are more affected by the environment for a longer period of time because both flowering and fruiting periods can be affected. In addition, predicting the harvest time and yield from a single seedling because multiple fruits can be harvested from a single seedling over time.

Therefore, research has been conducted in various facilities to analyze and predict factors related to fruit and vegetable yields. Previous studies have analyzed and predicted daily yields of tomatoes grown in a greenhouse [5] and a study that applied nonparametric regression to analyze and predict yields for tomatoes also grown in a greenhouse [7]. These studies targeted specific facilities and crops and have gone through a facility-dependent model generation process. However, since each facility has different equipment, environmental control methods, and production standards, developing a model that can be applied to all facilities and explain the factors in an integrated manner is difficult. In addition, it is difficult to generate a universal model because of changes in supply and demand due to social conditions, institutional reforms in agriculture, and the introduction of new equipment in factories.

Therefore, the purpose of this study is to identify the factors related to the yield of fruits and vegetables according to facility-specific conditions and to develop a model generation process that is applicable to various facilities. Here, the ultimate goal is to develop a yield prediction system that can be used by plant employees. The model generation process comprises (1) data selection, (2) data visualization, (3) feature design, (4) selection of prediction methods, and (5) model optimization.

## 2 RELATED RESEARCH

### 2.1 Prediction methods

Many studies have investigated yield prediction using various machine learning techniques, such as artificial neural networks and boosted regression trees [8][9][10]. Such techniques have achieved high prediction accuracy; however, machine learning is limited in its availability because it requires a large amount of training data.

In addition, to machine learning, statistical modeling has also been used as a prediction methods. Related studies include those by Hoshi et al. [5] and Okuno et al. [11]. In 2000, Hoshi et al. predicted daily yields of tomatoes grown in a greenhouse using topology case-based modeling (TCBM) [12] and multiple regression analysis. TCMB was the most accurate, with an average absolute error of 26%. In 2018, Okuno et al. proposed combining machine learning methods and statistical modeling for asparagus yield prediction. Using a Bayesian network as a machine learning method and multiple regression analysis for statistical modeling, the authors report that they were able to generate a regression model that can be used effectively as a clear basis for the prediction results and estimate the yield by capturing the trend of increase and decrease. However, validation of the prediction accuracy for small amounts of data and various seasons has not been carried out. To the best of our knowledge, comparative studies using multiple methods have not been carried out.

## 2.2 Feature generation

In developing a model, considering what features should be incorporated is an important task. Many of the related studies mentioned previously incorporate features related to environmental data, such as temperature, humidity, light, and precipitation. To shape the environmental data, only the most basic statistics, such as average, maximum, and minimum values over a period of time, are used to generate the features. However, with this method, finding factors related to yield is difficult because of the small variation in features, and the accuracy of the model cannot be increased to the maximum.

## 2.3 Research tasks

As mentioned previously, the model generation process formulated in this study consists of five sub-processes, performed in the following order: (1) data selection for model generation, (2) data visualization, (3) feature design, (4) selection of prediction methods, and (5) model optimization. From the issues in the related research, the major challenges are designing the features and selecting the prediction method. Therefore, there are two research tasks in this study.

### (1) Selecting a method to build a predictive model

Preparing a large amount of data is difficult for plant factories because of the significant environmental and cultivar variation caused by the introduction of equipment. Therefore, it is necessary to select a prediction model construction method that can be used even with a relatively small number of data. In other cases, it is also necessary to select a method whereby the process by which the final prediction result is calculated is easily understood. It is important to develop models that enable farmers to understand the theory of prediction models because unconvincing prediction models are unacceptable to farmers.

### (2) Selecting the feature extraction method

The actual yield data are considered time series data. A defining characteristic of time series data is that the datasets are closely related; therefore, it is necessary to generate features that can express the relationship and find the relationship with the target variable. Consequently, it is necessary to shape various features, such as median, variance, standard deviation, Fourier transform, and autocorrelation coefficients.

## 3 PROPOSED METHOD

### 3.1 Approaches

The approaches to the research tasks described in Section 2.3 are as follows.

(1) Selection of a method to construct a predictive model that can provide predictions even with small amounts of data and can interpret the model structure.

(2) Selection of methods to extract various features from the time series data.

A more detailed description of these approaches is given in Section 3.2. The model generation process that incorporates each approach is described.

### 3.2 Model generation process

As mentioned previously, the model generation process involves five sub-processes. In this research, we formulate these processes as a simple arithmetic procedure to enable us to identify factors related to yield and to generate a model that can accurately predict yield according to various facility-specific conditions. For each of the five sub-processes, we use a specific fruit as the base case to generate the model.

#### (1) Data selection

The data should be continuously acquired and recorded at each facility, rather than being based on costly surveys or confidential information. Therefore, in this research, initially, we only use environmental data to generate the model. The environmental data includes external weather data and internal environmental data. We select environmental data because it is common for plant factories to acquire and record external and internal environmental data.

#### (2) Data visualization

Data visualization facilitates the analysis of relationships between objective variables and features and identifies outliers. Scatter plots can be used to visualize data. A scatter plot takes one feature on the x-axis and another on the y-axis and plots the dots at each data point. In this case, since there are more than three features, we use a scatter plot matrix (paired plot) to plot all possible feature combinations. In addition, to obtain a clearer picture of the correlations between the attributes, we use a heat map of the correlation matrix as well as the scatter plot diagram.

### (3) Feature design

Here, we provide details of the approach described in Section 3.1, “Selection of methods to extract various features from the time series data.” Feature design involves designing a method to generate features and designing a method to select features. In time-series data, observed values and time points are recorded, and it is necessary to generate features that capture the characteristics of the data order and forward and backward relationships. Autocorrelation, which is a feature of time series data, can be applied to generate features. Autocorrelation is the correlation between time series dataset and data that is shifted several time steps. A previous study [5], used autocorrelation to generate features [5], and indicated that the features contributed significantly to prediction accuracy. Therefore, in this study, we also generate features from past yields.

In addition, we use a Python package called tsfresh (Time Series Feature Extraction on the basis of Scalable Hypothesis tests) [13] to generate various time-series-specific features. Although feature generation depends on the number of data, it is possible to generate close to 800 features per attribute. Some basic features include mean, maximum, and minimum values. Several other complex features, such as peak number, median, variance, standard deviation, features using Fourier transform and autocorrelation coefficients, and time-reversal symmetry features, are also available. TSFRESH is used because it can easily generate a large number of time-series data-specific features.

Next, we describe three feature selection methods. The first method uses the SELECT\_FEATURES\_TSFRESH function. This function uses statistical hypothesis testing to select only features that are likely to be statistically significant. The second method is mutual information-based feature selection. This method calculates how similar the product of the simultaneous distributions  $P(X, Y)$  and the individual distributions  $P(X)P(Y)$  are feature  $X$  and feature  $Y$ . The third method uses the Variance Inflation Factor (VIF) to check whether multicollinearity is occurring. If multicollinearity is present, the feature is selected by deleting the corresponding variable.

### (4) Prediction method selection

Here, the use of methods with a small amount of data and interpretable model structure are described. The selection criteria for the prediction method are: (i) It is possible to search for and evaluate the regression structure of the features for the prediction model. (ii) It is possible to construct a model with high prediction accuracy with a small number of data. We select the following three candidates for the modeling methods that can satisfy both criteria.

The first candidate is Multiple Linear Regression (MLR), which is a general statistical method for predicting continuous values of objective variables using two or more features. MLR is widely used to predict yields of various crops and also in fruit and vegetable yield prediction[5][9][11]. MLR is also selected because it has obtained some accuracy in prediction. To examine the best feature combinations, a stepwise method based on the Akaike Information Criterion (AIC), a common statistical variable

selection method, is implemented. The second candidate is the Generalized Additive Model (GAM), which is selected because it can provide predictions at the same level of accuracy as a machine learning model and retain the advantage of a linear model where the relationship between the objective variables and the features is easy to understand. Here, similar to MLR in the GAM, we also implemented the stepwise method based on the AIC. Finally, Multivariate Adaptive Regression Splines (MARS) were selected as the third candidate because, compared to GLM, it can explicitly represent the interaction between, features including tipping points in the tree structure [14].

### (4) Model optimization

We divide the target data in training data (75%) and test data (25%) and evaluate the datasets using the hold-out method. The correlation coefficient ( $R$ ) and the mean absolute error (MAE) are used as performance indicators of the regression model, where  $R$  measures the linear relationship between the predicted value and the measured value, and MAE is the average value (in physical units) of the difference between the predicted values. Since the percentage of yield varies from crop to crop, MAE is expressed as a percentage relative to the average yield.

For the developed model, the prediction accuracy is evaluated by  $R$  and MAE, and the most accurate prediction method is selected.

## 4 EXPERIMENTS

### 4.1 Target facility

The experimental facility in this study is a solar-powered plant factory located in Hakodate, Hokkaido (hereafter referred to as “Plant Factory A”). Plant Factory A owns two greenhouses that produce and sell hydroponically grown fruits and vegetables (7 fruits and 17 leafy vegetables). In addition to sensing environmental data, such as temperature, humidity, CO<sub>2</sub> concentration, and nutrient concentration in the hydroponic solution, Plant Factory A also collects external weather data, such as temperature, humidity, precipitation, and light intensity. Sensing both external and internal environmental data is useful for environmental control in the facilities.

We confirmed the demand for yield prediction of fruits and vegetables in Plant Factory A from interviews with employees in charge of management and employees in charge of actual production.

### 4.2 Target crop

We selected mini cucumbers as the crop for which the prediction model is developed. Cucumbers were selected as the base case because they are one of the major crops in the intake ranking of Japanese people [15]. Figure 1 shows a mini cucumber in actual cultivation. Mini cucumbers are the main crop cultivated at Plant Factory A. Plant Factory A has a large mini cucumber cultivation area. The product has a fast harvest cycle, and mini cucumbers are shipped almost every day. Therefore, we considered that there is a high demand for



growers to develop a yield prediction model for mini cucumbers.



Figure 1 Mini cucumbers being grown

### 4.3 Target data and problem setting

In this section, we describe the data used for model construction and the problem setting.

#### (1) Target data

The data used for model construction and evaluation were collected over a period of 67 weeks from February 1, 2018 to June 31, 2019. Plant Factory A measures and records internal environmental data and external meteorological data. Data are collected using sensors installed at several locations inside and outside the facilities. Data are recorded every minute, and average values are measured every 10 minutes. Table 1 lists the environmental data being measured. Nine environmental data are candidates for the features.

Table 1: Measured environmental data

Internal data	External data
<ul style="list-style-type: none"> <li>● Temperature in the house</li> <li>● House humidity</li> <li>● CO2 concentration</li> <li>● Satiation</li> <li>● Nutritive solution concentration</li> <li>● Illuminance</li> </ul>	<ul style="list-style-type: none"> <li>● Air temperature</li> <li>● Air humidity</li> <li>● Amount of light</li> </ul>

Next, interviews with producers were conducted as a reference for feature selection. Based on the interviews, we excluded nutrient concentration and illuminance from the features. Nutrient concentration was excluded because concentration of the nutrient solution was always constant and no significant change was observed when the variance of the data was actually checked. Illuminance was excluded because the irradiation time was defined as 15 hours from 4:00 am to 7:00 pm and no difference was observed between hours or days. We judged that even if we incorporated very small changes in the data into the features, the importance of the features in the constructed model would be very low, and there was a high possibility that they would introduce noise.

#### (2) Problem setting

By interviewing management and production employees, we set the yield for a single week from the day following the day on which the yield is predicted (hereafter, the prediction date) as the objective variable. The prediction date was fixed as every Saturday, and the dataset was created accordingly. The reason for setting every Saturday is that Plant Factory A has a meeting every Saturday to make a sales plan for the

following week and wants to use the data for the next week's yield forecast in that meeting. The objective variable, i.e., the yield for one week from the day following the prediction date (hereafter referred to as the weekly yield), is to be predicted using environmental and production data up to the forecast date.

### 4.4 Data visualization

#### (1) Visualization of Weekly Yield Trends

We analyzed the trends in weekly yields and the factors that contributed to the increase or decrease in yields. As a preprocessing step, we changed the time axis to weekly yields because the objective variable, yield, was classified in detail by plots and sowing dates. Figure 2 plots the changes in yields that shows the yield change in the period used for training the model. The average weekly yield was 135 kg (4500 plants), and the average daily yield was 19.3 kg (642 plants). These values are also used in the evaluation of the model.

First, we analyzed the trend patterns that appear in the time series data. Here, yield data is a type of time series data. Four main trend patterns appear in the time series data: trend variation, cyclical variation, seasonal variation, and irregular variation. The analysis shows that it is a trend pattern of irregular variation. Therefore, it is speculated that a method that can represent nonlinear behavior is likely to be appropriate as a prediction method.

Next, we identified periods when yields fluctuated significantly and analyzed the reasons for the change. The greatest increase in yield was in the period from late May to early October 2018. The number of plants was not a factor because the number of plants did not change significantly, i.e., it did not vary from season to season.

Based on these results, we decided to exclude the date data (harvest, sowing, and planting dates) and the number of plants from the features because there was no relationship with the yield.

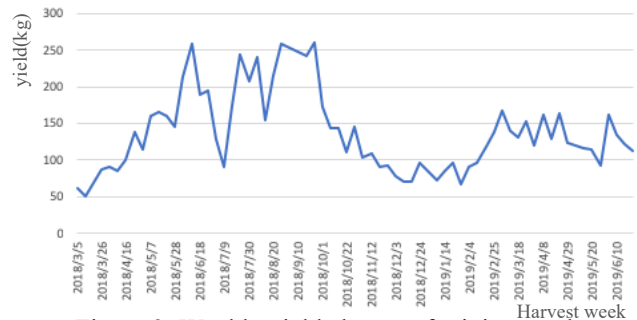


Figure 2: Weekly yield change of mini cucumber

#### (2) Visualization of relationships between attributes

The results of the analysis of the relationship between the objective variables and features and the relationship between the features are described. Figure 3 shows the scatter plot matrix, and Figure 4 shows the heat map table of the correlation coefficients. In Figure 3, the scatter plot matrix is an extract of the attributes that showed a particularly strong correlation with the objective attributes. Spearman's rank correlation coefficient was used to calculate the single correlation coefficient.



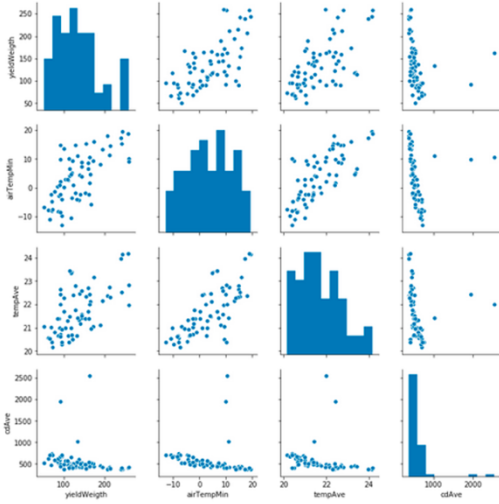


Figure: 3 Scatter plot matrix

Ten attributes, including the objective variables, were used in the analysis, and light preprocessing was done to facilitate the analysis. This section describes the preprocessing performed. First, the time axis of the recorded data was converted from every 10 minutes to days. Next, since several missing values were found, e.g., due sensing equipment malfunctions, the data were supplemented with the average values of the three days before and after the missing values. Finally, the period of the environmental data for the attributes was created using the seven days prior to the base date (the day before the prediction date). Seven days is the number of days from mini cucumbers flowering to harvest. Many studies report that the environmental factors in this period are important [9][16]. The list of attributes created as a result of the preprocessing is shown in Table 2.

Table 2: Potential predictor attributes in crop datasets

Attribute name	Attribute description
yieldWeight	Yield weight (kg)
airTempMax	Maximum temperature (°C)
airTempMin	Minimum temperature (°C)
airTempAve	Average temperature (°C)
outHumidAve	Humid (%)
lightIntensityAve	Solar radiation (kWh m <sup>-2</sup> )
tempAve	Average Temperature in the facility (°C)
tempMax	Maximum Temperature in the facility (°C)
tempMin	Minimum Temperature in the facility (°C)
cdAve	carbon dioxide concentration (ppm)

The strongest correlation among the environmental data was found for the maximum temperature with a single correlation coefficient of 0.78, followed by carbon dioxide concentration (−0.74) and mean temperature (0.68). It can be confirmed from the data that

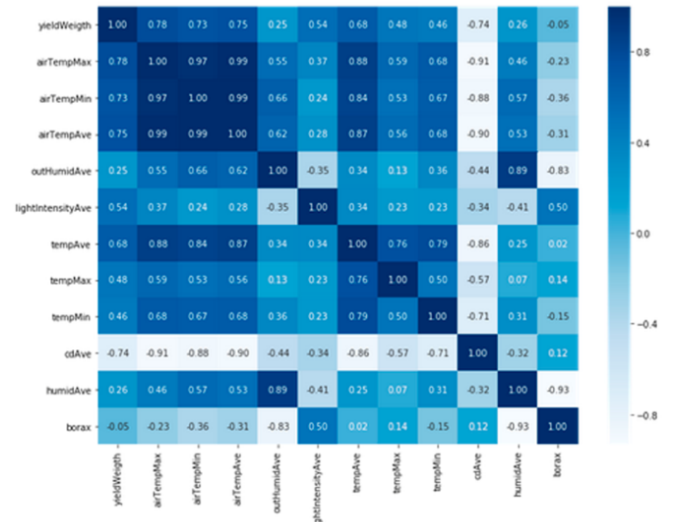


Figure:4 The heat map table of the correlation coefficients

the solar plant factory is affected by the external weather conditions in many ways. The reason for the strong correlation between maximum temperatures is that the optimum growing temperature for mini cucumbers is said to be approximately 18 °C–25 °C, and the growth is inhibited if the temperature is too high or too low. The paired plot diagrams showed outliers for some attributes, and it is necessary to consider how to process these data during preprocessing.

#### 4.5 Feature design

First, a total of 9802 features were extracted using TSFRESH. Features that cannot be handled, such as missing values and infinity, were removed. Next, the features were reduced to 204 using the statistical hypothesis test of the SELECT\_FEATURES function, which is also provided by TSFRESH. Finally, the three features were extracted by mutual information and the VIF. The features obtained as a result of the selection are shown in Table 3. We were able to extract time-series data-specific features. Model optimization and selection of the best prediction method were performed using these features.

#### 4.6 Model optimization

First, MLR, the GAM, and MARS methods were used to find the optimal feature combinations and hyperparameters in the training data. Table 4 summarizes the results of the most accurate feature combinations and the optimal hyperparameters. As can be seen the importance of the features differs for each method.

Table3: Scatter plot matrix

Feature code name	Feature name	Feature description
PYW	pastYieldWeight_cwt_coefficients__widths_(2, 5, 10, 20)_coeff_0_w_5	Previous week's yield of the objective variable, Calculates a continuous wavelet transform for the Ricker wavelet, also known as the Mexican hat wavelet.
LIA	lightIntensityAve_quantile__q_0.8'	solar radiation, Calculates the 0.8 quantile of lightIntensityAve.
ATA	airTempAve_cwt_coefficients__widths_(2, 5, 10, 20)_coeff_0_w_10' 41	Average temperature. Calculates a continuous wavelet transform for the Ricker wavelet.

Table 4: Model tuning results

Technique	Feature	Hyperparameter
MLR	PYW, LIA, ATA	
GAM	PYW, LIA, ATA	Select = FALSE Method = REML
MARS	PYW	degree=3 nprune=2

Table 5 shows the predicted accuracy results in the training and test data using the tuned features and hyperparameters. From the evaluation results of each accuracy measure, it was found that, among the three methods, MARS had the highest prediction accuracy for the test data. Overall, there was a pronounced tendency to overlearn. Figure 5 is a graph of the measured and predicted weekly yields.

Table5: Evaluation results for each accuracy index

Technique	R		MAE (%)	
	Train	Test	Train	Test
MLR	0.875	-0.324	16	51.9
GAM	0.898	-0.340	14.6	54.9
MARS	0.820	0.279	20.0	21.0

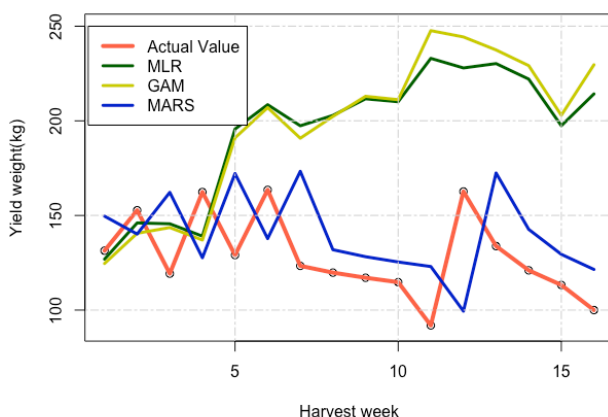


Figure 5: Measured and predicted results of weekly yields

## 4.7 Discussion

First, we discuss the results of the feature selection for each method. PYW and LIA (Table 4) were selected for all three methods. PYW refers to past yields and is an autocorrelation feature specific to time series data. The LIA is the amount of light and is affected by the weather, which is a characteristic of a solar plant factory, which was confirmed from the model structure.

Next, we consider the results for the test data for the MARS model, which had the highest accuracy. The  $R = 0.279$  value (Table 5) indicates that there is a weak correlation between the measured and predicted values. Compared to a previous related study [9], which investigated the accuracy of prediction models using multiple machine learning methods,  $R = 0.42$  (the model with the highest  $R$  value), is low accuracy. It can be seen from Figure 5 that the model does not accurately represent the timing of large increases and decreases in yield. The results show that there is still room for improvement with regard to accuracy. In this study, we made predictions using only features related to environmental data; however, we believe that there is a limit to the accuracy of predictions using environmental data alone. A study by Hoshi et al [5] reported that production engineering factors, such as working hours and days of the week, rather than the air temperature and solar radiation, were more than twice related to daily production with a correlation coefficient. Since various factors, such as environmental and production engineering factors, are involved in plant factories, it is necessary to investigate new features in the future. On the other hand, the MAE was 21.0%, which is almost equivalent to the prediction accuracy of 18.12% in a previous study [9]. This prediction accuracy is comparable to that of machine learning.

We asked the employees of the plant factory to confirm the results of this study, and they said that the accuracy of this study was not yet at the practical level. They pointed out that to achieve a practical level, it is more important to know the trend of yield increase or decrease than the error. In the future, we will improve the accuracy with the goal of capturing yield trends to meet the needs of employees in the plant factory.

The present experiments suggest that MARS can be used to develop a model that shows a certain prediction accuracy. However, there is room for improvement in terms of errors.

## 5 CONCLUSION

## 5.1 Summary

The purpose of this study was to identify the factors related to the yield of fruits and vegetables according to facility-specific conditions and to develop a model generation process to predict the yield that is applicable to various facilities. To address the two research problems, (1) examination of methods to extract features and (2) examination of methods to construct a prediction model, we have adopted the following approaches: (1) selection of a method to construct a predictive model that can perform effectively even with small amounts of data and can interpret the model structure, and (2) selection of methods to extract various features from the features of time series data. We generated new features from the acquired data and searched for the optimal hyperparameters. The experimental results demonstrate that the importance of the features differs for each statistical modeling method. Among the three statistical modeling methods, the model using MARS provided the most accurate yield predictions.

## 5.2 Future tasks

- (1) Verification of the applicability of the model generation process to crops in other environments

In this paper, an optimal model generation process was developed using the mini cucumber as the base case. In future, we will evaluate the usefulness of the proposed method by verifying whether the developed process can be applied in different environments.

- (2) Development of a Yield Prediction System

We will develop a web system that displays the results of a week's yield prediction based on the various prediction models we have developed. The users of the system are the management and production employees of the plant factory. It is envisioned that yield prediction system will help system users to develop accurate production and shipping plans.

## REFERENCES

- [1] T. Iike, "NAPA Research Report 2018, Chapter 3: Plant Factory Business - Current Status and Future Prospects of Plant Factory Management" <[https://www.nomuraholdings.com/jp/company/group/napa/data/20180219\\_03.pdf](https://www.nomuraholdings.com/jp/company/group/napa/data/20180219_03.pdf)> [Accessed December 12, 2019](in Japanese).
- [2] Japan Greenhouse Horticulture Association, "Large-scale horticulture and plant factory survey and case studies" <<https://www.maff.go.jp/j/seisan/ryutu/engei/sisetsu/pdf/daikibo.pdf>> [Accessed December 12, 2019](in Japanese).
- [3] Yano Research Institute, "Conducted a study on the "Next Generation Plant Plant Factory" market with high functionality and high added value (2018)" <[https://www.yano.co.jp/press-release/show/press\\_id/1990](https://www.yano.co.jp/press-release/show/press_id/1990)> [Accessed December 12, 2019](in Japanese).
- [4] Ministry of Agriculture, Forestry and Fisheries, "Explanation of the Plant Factory, Japan Center for Social Development Research" <<http://www.maff.go.jp/j/heyasodan/1308/01.html>> [Accessed December 12, 2019](in Japanese).
- [5] T. Hoshi, T. Sasaki, and H. Tsutsui, "A daily harvest prediction model of cherry tomatoes by mining from past averaging data and using topological case-based modeling", Computers and Electronics in Agriculture, Vol. 29, No. 1, pp. 149-160 (2000).
- [6] T. Kozai, "Plant Factory Basics in Illustrated Manuals: From Capital Investment and Production Costs to Hydroponics Technology, Distribution, Sales and Management", SEIBUNDO SHINKOSHA Publishing Co., LTD (2010) [Accessed December 12, 2019](in Japanese).
- [7] M. Oucouchi, M. Aono, K. Kwashima, Estimating production of greenhouse crops by using Nonparametric Regressions, DEIM Forum, E5-2(2011) (in Japanese).
- [8] A. Chlingaryan, S. Sukkarieh, and B. Whelan: Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review, Comput. Electron. Agric., Vol. 151, pp. 61–69(2018).
- [9] González-Sánchez, Alberto Frausto-Solis, Juan and Ojeda, Waldo, Predictive ability of machine learning methods for massive crop yield prediction, SPANISH JOURNAL OF AGRICULTURAL RESEARCH (2014).
- [10] K. Kuwata, and R. Shibasaki, Estimating crop yields with deep learning and remotely sensed data, 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, pp. 858-861(2015).
- [11] G. Okuno, S. Niiya, Yield estimation of asparagus using a combination of machine learning and statistical modeling, Japan Social Data Science Society, Vol. 2, No. 1, pp. 14-18(2018) (in Japanese).
- [12] H. Tutui, Z. Kurosaki, T. Sato, Nonlinear Modeling Technique Using Historical Data for Case TCBM: Topological Case Based Modeling, Transactions of the Society of Instrument and Control Engineers, Vol. 33, No. 9, pp. 947–954(1997) (in Japanese).
- [13] Christ, M., Braun, N., Neuffer, J. and Kempa-Liehr A.W. (2018). Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh -- A Python package). Neurocomputing, Vol. 307, pp. 72-77(2018).
- [14] T. Matsui, T. Ugada, N. Machimura, Analysis of Factors of Abandonment and Development of Prediction Model by Machine Learning Algorithm, Journal of Japan Society of Civil Engineers, Vol. 70, No. 6, pp. II\_131-II\_139(2014) (in Japanese).
- [15] Ministry of Health, Labour and Welfare, "Vegetable Intake Ranking in Japan" <<https://www.mhlw.go.jp/file/04-Houdouhappyou-10904750-Kenkoukyoku-Gantaisakukenkouzoushinka/0000096137.pdf>> [Accessed December 12, 2019] (in Japanese).
- [16] H. Itoh, R. Yamashita, An Analysis of Lettuce Growth in a Plant Factory Multivariate Analysis on Time Course of Fresh Weight, Vol. 11, No. 1, pp. 50-58(1999) (in Japanese).



# Proposal for Method of Efficient Building Facility Management linking Human Sense and Information of BEMS

Fumiaki Kimura\*, Fuyuki Sato\*, Yoshihito Endo\*, Takeyuki Kimura\*, and Shinji Kitagami\*\*

\*Mitsubishi Electric Building Techno Service Co., Ltd.

\*\*Fukui University of Technology, Japan

{kimura.fumiaki}@meltec.co.jp

**Abstract** - The decrease in working population has become a social problem because of the decreasing birthrate and aging population in Japan. In the building facilities management industry, as the number of large office buildings has increased since the collapse of the “bubble economy,” the sense of human resources shortages in management and administration is increasing. Furthermore, the building facility management business is a unique service based on operational experience in diversified office buildings. Although we have accumulated know-how in the operations of large-scale office buildings, transferring know-how is becoming a challenge as the population ages more than ever.

We propose a method to improve the efficiency of facility management operations and the transfer of know-how using the information from building tenants, which was not previously utilized. In this study, we used text mining and machine learning methods to convert the daily report data, including inquiries and complaints from tenants into relevant information for facility managers. We also performed a simulation using the daily data of large office buildings in operation to evaluate our proposal.

**Keywords:** Building Facilities Management, BEMS, text mining, machine learning, Large Office Buildings

## 1 INTRODUCTION

The decrease in the working population has become a social problem due to decreasing birthrate and aging of the population in Japan. In the building facilities management industry, as the number of large office buildings have increased since the collapse of the “bubble economy,” the sense of human resources shortages in management and administration is increasing. In addition, the building facility management business is a unique service based on operational experience in diversified office buildings. Although we have accumulated know-how in the operations of large-scale office buildings, transferring know-how is becoming a challenge as the population ages more than ever. [1] [2].

In recent years, large office buildings adjacent to commercial facilities have been attracting more users even on weekends and holidays. Moreover, many buildings with foreign companies as tenants operate at night because of the time difference with their home countries. In such a large office building, as users of the building are becoming more diverse, the building facilities management service is

performed by the facilities staff 24 hours a day, 365 days a year, and staff is permanently assigned to manage and operate the facilities.

They record all results of periodic inspections, response to inquiries and complaints, and Building and Energy Management System (BEMS) alarm response, which is carried out as part of the facilities management, and report to the building management in a daily report format [3]. The building manager builds effective relationships with tenant contractors based on these reports. Although these reports are useful to reduce the number of inquiries and complaints from tenants and other users, they seldom use it to improve facility management operations.

Since there is concern that the shortage of human resources for facility management will continue to become more serious in the future, the need arises to utilize the information collected from users of large office buildings to improve the efficiency of facilities management operations and to transfer know-how.

In this paper, we propose a method to improve the efficiency of facility management operations and know-how transfers using the information obtained from tenants in the building, which was not previously used. We used text mining and machine learning methods to convert the daily report data, including inquiries and complaints from tenants into relevant information for facility managers. In addition, we performed a simulation using the daily data of large office buildings in operation to evaluate our proposal.

## 2 BUILDING FACILITIES MANAGEMENT

### 2.1 Overview of building facilities management services

In general, maintenance and management of a building are outsourced to specialized companies. Outsourcing is categorized into three areas: cleaning and sanitation management, facility management, security and disaster prevention. Facility management includes the general management of electrical equipment, air-conditioning equipment, water supply, sanitation, and drainage systems to detect abnormality early and respond to emergencies. Information on facility malfunctions can be obtained by the BEMS installed in the building as an alarm.

On the other hand, tenants in the building make inquiries by telephone to the facility manager if they detect any abnormality in the facility. Upon receiving these inquiries, the facility manager investigates the cause and takes corrective action. In addition, they carry out regular inspections according to preset inspection items to check for the normality of the facility. All work performed by these facility managers is recorded in a daily report and reported to the owner. In this way, facility management works are diverse, and the required knowledge is extensive.

## 2.2 Issues in building facilities management

In the facilities management of a building, the building manager receives alarm warnings detected by BEMS and inquiries from tenants and records the details of these operations after responding to them. They are reported to the building owner in the form of daily, monthly, and annual reports. However, these routinely accumulated data are rarely used for anything other than reporting to the building owner. Ideally, it is desirable to analyze the recorded information of responses to improve the efficiency of operations, not only to reduce the number of facility malfunctions and inquiries from tenant users [4][5].

The following are the issues facing building facilities management nowadays:

- (1) The reason for the lack of progress in the analysis of daily report data is related to the characteristics of building facilities management operations. In other words, because there are many types of facilities and the manufacturers and models of equipment installed in each building are different, facility management work requires more extensive knowledge than expertise. For facilities that require a high level of expertise, a method to code and analyze the causes of failures has been reported [6]. However, it is difficult to utilize the data for analysis because facility management for various facilities accumulates data in natural language.
- (2) In the facilities management of a large-scale building,

several people work in shifts at night and on holidays. Therefore, the experience gained in responding to inquiries and alarms is accumulated by the individual. In other words, in the event that the facility manager on duty on the same day receives an inquiry that he has never experienced before, or an alarm from BEMS, the manager may have to respond to it without having any relevant information. In this situation of urgent response, the facility manager has to deal with the lack of information, and there is a problem of repeating inefficient work, such as return to pick up the necessary tools repeatedly or doing unnecessary work.

## 3 PROPOSAL METHOD

To solve the problems of building facilities management described above, we have developed a new system for building facilities management. In this chapter, we propose a method to enable the know-how transfer and efficiency of facility management by utilizing the accumulated daily report data. The proposed method consists of a function to analyze daily data and a function to utilize the analyzed data.

### 3.1 Outline of the proposal

Figure 1 shows the concept of a proposed method that enables the transmission of know-how and efficiency in building facilities management operations. The proposed method consists of a daily data analysis function and a data utilization function.

The daily data analysis function enables the facility manager to input the results of alarms from the BEMS, inquiries from building users, and responses into the proposed system in the same way as in the past. Since the input data are created in natural language, they are processed in a form that can be analyzed. The processing methods include text mining [7][8], which involves word-by-word segmentation, cleansing to prevent duplication of synonyms and similar words, and machine learning to classify them into forms that can be analyzed [9][10].

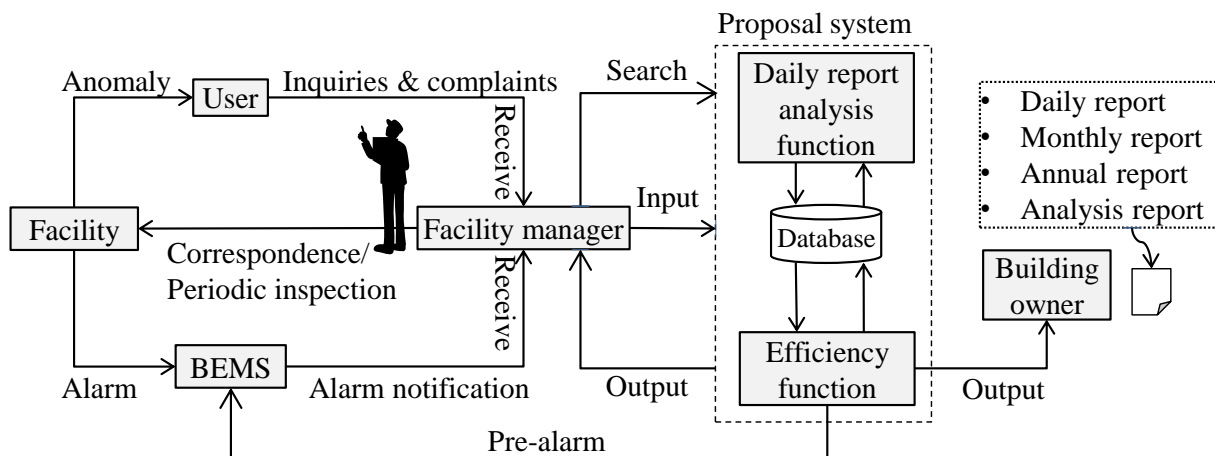


Figure 1: Proposed method concept



In addition, the data utilization function uses the data accumulated in the database in a form that can be analyzed and outputs information to support the facility manager's know-how when responding to alarms from BEMS and inquiries from users. At the time of the periodic inspection, the information is output as information on the transfer of know-how for the predetermined inspection items. Additionally, it is output as a daily report to be reported to the owner. We extract the inquiries and alarms information that occur at regular intervals from the analyzed data and input this information to BEMS. When the preset conditions are met, BEMS outputs a pre-alarm to enable us to respond systematically before an alarm or inquiry occurs. Thus, the proposed method utilizes the accumulated information and supports the work of facility managers.

### 3.2 Daily report data analysis function

As illustrated in Figure 2, this function analyzes the data recorded by the facility manager in natural text when an alarm from BEMS or a query from the user is received and when the response is completed, and the data is processed into a form that is later utilized. Analyze the data and process it into the usable form. This method performs natural language processing of recorded data using text mining and machine learning with rule-based and teacher data. The results of the analysis are then used to create data for search and visualization.

As shown in Figure 3, this text mining process divides natural sentences entered by the facility manager into the smallest meaningful words using a morphological analysis method. There are a few building-specific proper nouns, such as tenant names and facility names, among others, that are divided into these words. To this end, we register tenant names and technical terms used in facility management in advance, so that the function can recognize them in the dictionary. Since there are often many synonyms and similar words that interfere with natural language processing, we perform cleansing and extract only the words needed for analysis. As shown in Figure 4, the words created by text mining are automatically classified in machine learning. Six items (equipment to be queried, equipment category, cause, queried category, symptom, and treatment) are classified using a commercially available machine learning software. To increase the accuracy of classification, information for classification is registered in the teacher data. The data output from this machine learning is stored in the database and used in 3.3 Data Utilization Function.

Text mining is a method of extracting useful information from text data by dividing the text into words (nouns, verbs, adjectives, etc.) and analyzing their frequency of occurrence and correlations using natural language analysis techniques. The proposed method is used to extract the words necessary for the analysis of the equipment. In addition, cleansing is the role of removing such data which can cause problems such as counting many synonyms and similar words when they are included in more than one place. In this proposal, we use cleansing to define such data as referring to the same floor in a building, since a building may show more than one indication of a location, for example, 20th floor, 20th

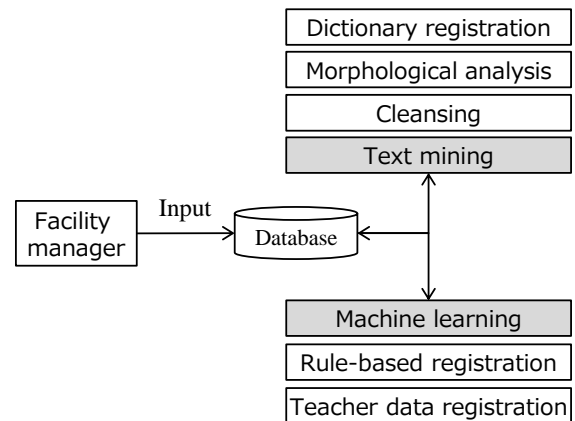


Figure 2: Daily report analysis function block diagram

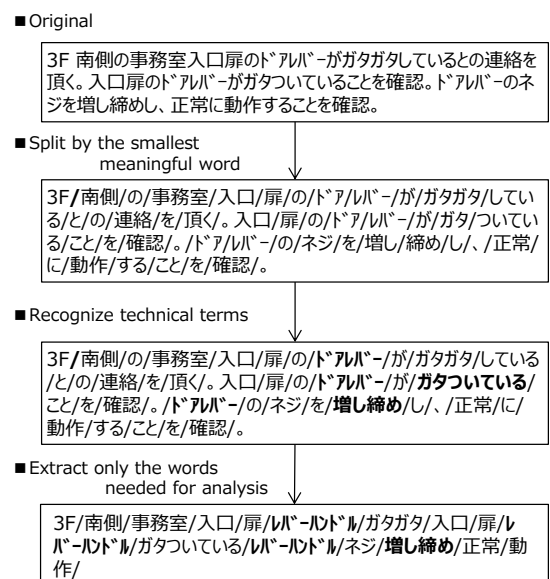


Figure 3: Text mining block diagram

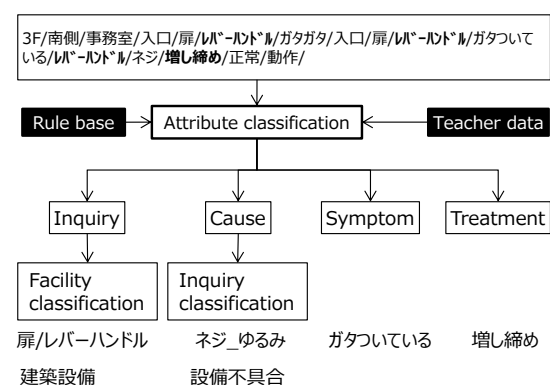


Figure 4: Machine learning block diagram

level, twentieth floor, etc., which is recognized as a different floor, so we use cleansing to define such data as referring to the same floor.

### 3.3 Data utilization function

The daily report data analysis function allows us to receive alarms from BEMS, user inquiries, and responses to them. The results are processed and stored in a form that can be analyzed. Based on the data, the facility manager enters the received data and performs a search to provide the necessary information to the facility manager. The search and result display screens are shown in Figure 5. However, with this method, we found that it was difficult in many cases to retrieve the history of characters in daily reports. Therefore, we propose to use the text mining process, which is shown in the function to analyze the data processed in 3.2 and machine learning for word extraction, cleansing, and machine learning attribute classification for analysis. This enables the facility manager to respond to the results of a search that does not yield the information needed in a traditional search, rather than the results of a search. We can provide the information necessary to perform. As an example of the search results, we cured the leakage by bringing plastic sheeting to protect the electrical equipment racks and wires. This information is expected to enable even inexperienced facility managers who lack the know-how to respond based on the same information as experienced facility managers. We believe that this function solves the problem of not being able to make use of the information that has been accumulated from user inquiries and the information resulting from responding to those inquiries, and, improves the efficiency of the facility manager by making use of the data.

The duties of the facility manager include periodic inspections of facilities, which are planned and carried out. The checks are carried out according to the inspection items prepared in advance. However, the same person does not always inspect the same equipment. The facilities management work is done in shifts, including night shifts. As a result, there are differences in information and time between operators. Therefore, by utilizing the data accumulated in 3.2, we have created an inspection item in advance and added the know-how information to it. Additions are made. With this added information, it is possible to work after understanding the focal points. Furthermore, the efficiency of the information terminals may be impaired due to the increase in workload, so that the information terminals can be used for the same purpose. A mechanism to check the items after they are completed is added to the system. This checking mechanism allows you to see the difference in work time and makes the difference in work time for the same work item clear. It is possible to compare with other workers by using graphs, etc., and to get know-how from workers who have less time to work. It is possible to share the. This can be expected to improve the efficiency of building equipment management operations by passing on the work know-how.

The method for obtaining information by visualization is shown in Figure 6. It visualizes on which floor the inquiries are occurring. In addition, by selecting the corresponding floor, the floor plan is displayed, and information such as which section of the floor generates the most inquiries is visualized by using a bar graph. This visualization enables

the facility manager to search for inquiries from tenants on the search screen shown in Figure 5, and to obtain



Figure 5: Machine learning block diagram

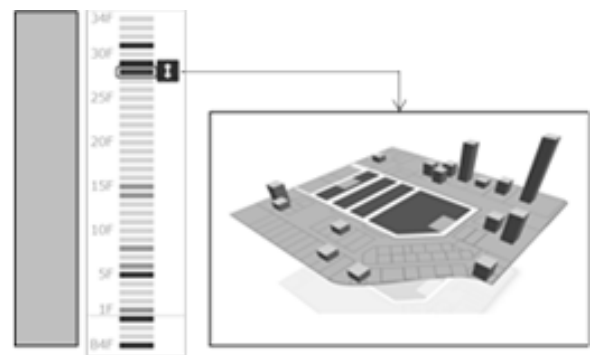


Figure 6: Visualization of inquiries and alarms

information such as what kind of inquiries were received in the past, the results of responses, the tools required for responses, and the causes identified in the past. More still, the periodic inspection enables us to understand the tenant's tendency in advance, therefore we can understand the points to be paid attention to in advance. In this way, we believe that this function can also be used to visualize the characteristics of individual tenants, such as the kind of inquiries they send out. The plan view in Figure 6 can be created by using the Computer Aided Design (CAD) data at the time of delivery of the BEMS.

Moreover, as a result of analyzing the data through visualization, we were able to obtain alarms from the BEMS and the data from users. It may turn out to be a periodic occurrence in inquiries. In this case, the Building Automation and Control Networking Protocol (BACnet) communication interface is used to respond to the BEMS dummy alarm signal. Systematic additions to the items of the periodic inspection by generating a pre-alarm and taking action makes it possible to prevent alarms and inquiries from occurring.

## 4 SIMULATION AND EVALUATION

In this study, using data of two large office buildings accumulated over the past ten years: 3.2 Daily report data analysis function and 3.3 Data utilization function provide useful information to facility managers. We evaluated the functions provided by the simulation.

### 4.1 Daily report data analysis function

The number of response data accumulated over the past ten years in the buildings in question is approximately 20,000. The most recent of these data is split and inputted, and then the teacher data is modified to match the input, and Simulations were performed on the number of data that were stable with machine learning accuracy above 90%. The simulation results are shown in Figure 7. As a result of 333 inquiries from tenants on November 1, the machine learning accuracy was 95%. The effective data rate of the teacher data was 71.9%. Then, on January 12, 8,909 entries were recorded, with a machine learning accuracy of 90.0% and a teacher data. The valid data rate was 95.0%. Subsequently, as of March 7, there were inquiries, the accuracy of machine learning was 96%, and the effective data rate of the teacher data was 100%. Since then, the accuracy of machine learning has mostly been unchanged. The results show the accuracy of machine learning fluctuates after data input and the effective data rate of the teacher data. The amplitude range is reduced, and the effective data rate of the teacher data is 100%, and the accuracy of machine learning is 95%. The data are considered usable when they exceed the value.

### 4.2 Data utilization function

The results of the analysis by the daily data analysis function were used to simulate a form of information that can be effectively utilized by facility managers. The date, time, place, content, cause, treatment, and facilities of the inquiries and alarms are extracted from the daily report data, which is the history of trouble response. In this way, it is possible to reduce the number of problems by predicting the occurrence of problems at the right time before a query occurs and dealing with them by systematic inspection work.

A comparison of the number of inquiries from different tenants on different floors during the same period showed that there were differences in the content of the inquiries. The results are shown in Figures 8 and 9. These results show differences in the points of concern for each tenant. Tenant A made 702 inquiries in six months, as shown in Figure 8. Of these, more than half or 452 requests for temperature changes were received, indicating that people are highly interested in temperature.

On the other hand, Tenant B made 179 inquiries in six months, as shown in Figure 9. Of these, 32 requests to change the temperature and 52 inquiries about door noise were received. In other words, Tenant A's inquiry about door noise was the third most common inquiry with 51 inquiries, but Tenant B was more interested in indoor noise. The number of inquiries about door hinges for the entire building is shown in Figure 10. From the analysis of the

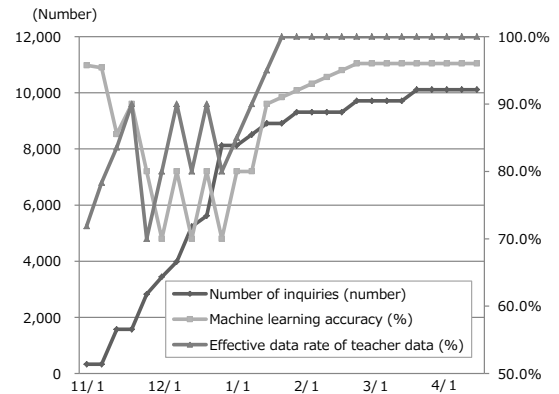


Figure 7: Visualization of inquiries and alarms

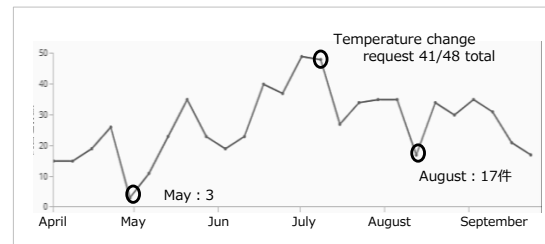


Figure 8: Tenant A Inquiries

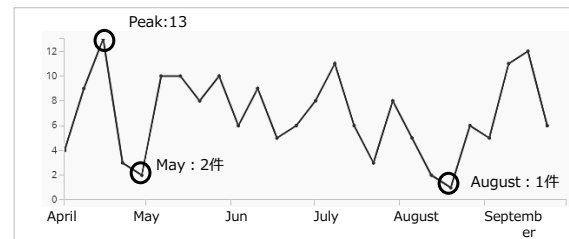


Figure 9: Tenant B Inquiries

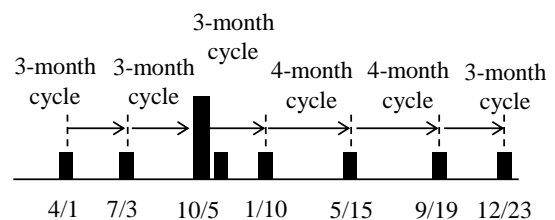


Figure 10: Door hinge inquiry interval

number of these inquiries, we can confirm that they occur in a constant cycle of about 3 to 4 months. If we can analyze these periodicities, we can prevent the inquiries in advance. In other words, if the periodicity can be detected from the analysis results, the information is registered in BEMS, and the pre-alarm function enables systematic work to be done before the inquiry occurs. Based on the results of these analyses, we believe that the efficiency of the facility manager's work can be improved by considering the characteristics of the tenants when they perform periodic inspections and responds to inquiries.

## 5 DISCUSSION

In this chapter, we evaluate the results of simulations based on the daily reports recorded in the facilities management in the two large office buildings described in Chapter 4.

## 5.1 Daily report data analysis function

Text mining was used to divide natural sentences into word units and register them into dictionaries to extract words used in facility management, about 500 words were registered for 1,000 daily reports. We started by organizing the word list since there have been no research results in the field of facilities management and no commercially available software. Although we spent much time registering this dictionary, it is necessary to consider how much the dictionary can be used when applied to buildings other than the target building.

In addition, we spent much time checking the accuracy of machine learning and adjusting the teacher data on a case-by-case basis while incrementally increasing the daily data. Since the daily data of the target large office building has about 100,000 records, we initially predicted that the accuracy of the data would be about 1% of that data, and the accuracy would remain approximately the same by adjusting the teacher data. However, the results show that up to 10% is necessary. This adjustment of teacher data, as well as dictionary registration in text mining needs to be considered with respect to its generality for use in other buildings. It is also expected to be closely related to dictionary registration in text mining. In the future, we need to continue to test the accuracy of machine learning with about 100,000 data points.

## 5.2 Data utilization function

Considering the function of data utilization by using the results of the analysis of accumulated daily data conducted in 4.1. Analysis of the results of the daily data analysis function shows that even for tenants on different floors of the same building, different results have been found in the queries. Though facility managers were aware of this, until now, no data existed that could be presented in the form of data. It is meaningful to use these results to understand the characteristics of each tenant and to respond to their inquiries. However, it is necessary to conduct many of these analyses and incorporate them into daily operations to examine the extent to which it is possible to reduce the response of facility managers. For this purpose, it is necessary to proceed with the analysis and study a system that can deal with problems at the appropriate time by generating a pre-alarm before BEMS function generates a query or alarm, without installing new sensors. Furthermore, it is necessary to discover the periodicity of the occurrence of inquiries, etc., from the data analysis with the pre-alarm function of BEMS, and to consider systematically carrying out inspections at an appropriate time by sending data to BEMS before an inquiry occurs, thus triggering a pre-alarm.

## 6 CONCLUSION

In this paper, we propose to use the daily report data accumulated by facility managers for text mining, and machine learning. The analysis data is prepared by using a computer program, and the results are made available in a form that allows the facility manager to make use of them. We worked to improve efficiency. In the future, we will build a function linked to BEMS in consideration of the workflow of the facility manager, and we will be able to provide human information to the facility manager. It is captured as a BEMS sensor, and the BEMS generates a pre-alarm for the user. The facility manager can respond to inquiries from the users utilizing inspections before they occur. We believe that it is necessary to consider how to achieve this goal. We will further study the systematic management method to improve the efficiency of facility management operations and to hand down the know-how. We believe that we need to do this.

## REFERENCES

- [1] Japan Building Maintenance Association, 48th fact-finding report [public version], building maintenance information yearbook, pp. 19 – 22 (2018).
- [2] Matsumoto Takuso, Management risk of building maintenance business due to lack of manpower: population decline is a major management risk, Equipment, and management, 49(8), No.5, pp. 42 – 51 (2015).
- [3] Takeshi Watanabe, Building management system in IoT Age, J. IEIE Jpn, Vol.37, No.3, pp. 180 – 183(2017).
- [4] Akira Takakusagi, Chiharu Nishi, Seiichi Honda, and Yukihide Koizumi, Setting method of quality level of jobs for contract in maintenance and management of building services, AIJ J. Technol. Des., Vol. 13 No. 25, pp. 197 – 202(2007).
- [5] Masayoshi Komatsu, and Akira Takakusagi, Study on defects and maintenance work at small and medium scale office buildings, J. Archit. Plann. AIJ, No. 574, pp. 161 – 168(2003).
- [6] Mari Nagasaka, Makoto Sato, and Eiji Kinoshita, Maintenance optimization with various facility data using Ontology processing and machine Learning, IEE Jpn, pp. 1422 – 1427(2017).
- [7] Shigeaki Sakurai, Akihiro Suyama, and Yumi Ichimura, The effect of a key concept dictionary for a text mining system, JSAI, pp. 1422 – 1427(2002).
- [8] Yumi Ichimura, Yasuko Nakayama, Toshio Akahane, Miyoko Miyoshi, Toshikazu Sekiguchi, and Yousuke Fujiwara, Text mining for Salesperson's daily reports: extraction of best practice and missed opportunity, JSAI, pp. 532 – 534(2000).
- [9] Yumi Ichimura, Akihiro Suyama, Shigeaki Sakurai, and Ryohei Orihara, Knowledge dictionary development tool, IPSJ research report, pp. 25 – 31(2001).
- [10] Kanji Takahashi, and Yuki Okuda, Assisting text classification for request post analysis, IFAT, Vol.133, pp. 1 – 6(2019).

# Regression Combined with Random Forest for Wagyu Quality Prediction

Shingo Tsukamoto<sup>†</sup>, Haruka Ikegami<sup>‡</sup>, Tamako Matsushashi<sup>‡</sup>, Kazuya Matsumoto<sup>‡</sup>, and Takuya Yoshihiro<sup>\*</sup>

<sup>†</sup>Graduate School of Systems Engineering, Wakayama University, Japan

<sup>‡</sup>Graduate School of Biology-Oriented Science and Technology, Kindai University, Japan

<sup>\*</sup> Faculty of Systems Engineering, Wakayama University, Japan  
s216169@wakayama-u.ac.jp

**Abstract** - As Wagyu has been recognized as high-quality beef in the world, livestock farmers of Wagyu are continuing their trials to improve beef quality of their cattle. However, since improvements have been done based on their own experiences and know-how, inheriting their techniques to others is hard. To establish efficient way of raising cattle based on scientific evidence and populating it would lead to high productivity of Wagyu beef. Wagyu beef is evaluated and priced based on quality. One of the most important quality criteria is carcass weight, which determines the amount of beef meat to be sold. In this study, we propose a new method to predict carcass weight in combination with the traditional linear regression and the random forest method. First, we use linear regression to predict carcass weight from the initial weight and protein expression dataset of cattle, and then learn its error from unused variables using a random forest. Through evaluation, we found the proposed method outperforms the conventional methods.

**Keywords:** Wagyu, Carcass Weight, Random Forest.

## 1 Introduction

Wagyu is a high quality foodstuff in Japan that is attracting attention from abroad, and there is a growing interest in increasing the value of Wagyu beef. The economic value of Wagyu is assessed by a variety of indicators, with six main traits: CW (Carcass Weight), BMS (Beef Marbling Standard), TE (Yield Enhancement), RT (Rib Thickness), SFT (Subcutaneous Fat Thickness) and REA (Rib Eye Area). Among them, CW is particularly important as it determines the total amount of Wagyu beef shipped.

It is believed that the traits of Wagyu are determined by two types of factors: genetic and environmental. Farmers can only change meat quality by manipulating two types of environmental factors when fattening purchased calves. Environmental factors affect the growth of beef cattle, and a variety of factors such as feed, climate, and physical activity during fattening affect it. Farmers are making various efforts every day to improve traits by manipulating environmental factors. However, each farmer's fattening methods are based on the knowledge and know-how obtained from their own experience, which makes it difficult to transfer or inherit the fattening methods. As a result, to produce a mass of stable beef cattle is not realistic. There is a need for science-based fattening methods to improve the productivity of beef cattle to solve this problem.

The purpose of this study is to predict CW of cattle when

slaughtering it based on the initial body weight when a breeding farmer bought the calf and the protein expression profiles measured during the fattening period. If the final CW value can be predicted during fattening period, breeding farmers would be able to improve their fattening method, and also read to establish efficient fattening methods based on scientific grounds.

In this study, we propose a method for predicting the CW of Wagyu in combination with the regression analysis and the Random Forest method. The purpose of this study is to predict CW from the initial weight of the cattle and protein expression profiles measured during the fattening period. It is known that there is a large correlation between CW and the initial weight of Wagyu. We aimed to build a prediction method that takes advantage of this correlation. In our method, we first make predictions using single regression. Then, by applying machine learning to the prediction errors in regression, the prediction errors are reduced. The Random Forest method was chosen as the machine learning method to be applied. The explanatory variable, analytical data, is considered to be very high-dimensional as it is obtained comprehensively. Random forests have features that make them less prone to overfitting and less likely to reduce estimation accuracy, even when the number of samples is small and the number of explanatory variables is large. This feature makes it possible to perform analyses that are less sensitive to the number of dimensions of the analysis data.

## 2 Basic Knowledge

### 2.1 Wagyu and Quality Assessment

Wagyu is a kind of cattle created by combining native Japanese cattle with foreign breeds, and consists of four breeds: Japanese Black, Japanese Brown, Japanese Shorthorn, and Japanese Polled. Wagyu beef has a tendency to be marbled, which is the characteristic that foreign beef does not have. In order to differentiate Wagyu from foreign beef, breed improvement has been promoted so that Wagyu beef has a good marbling characteristic with fat. Owing to the long effort to produce high quality marbling beef, now Wagyu beef is sold at high prices. Wagyu breeders have been continuing their effort to gain experience in breeding better quality cattle. The farmer raises the calves and ships them to a meat center. They are processed into carcasses at the meat center and auctioned off with the carcass performance information. The carcass is the body after removing the head, limbs, tail and internal organs. There are many indicators for quality of Wagyu beef, and

among them, there are six particularly important traits, that is, CW (Carcass Weight), BMS (Beef Marbling Standard), YE (Yield Enhancement), RT (Rib Thickness) and SFT (Subcutaneous Fat Thickness). They are referred to as the six major traits. Among the six major traits, CW is considered as a particularly important indicator because it represents the total amount of Wagyu beef shipped. Wagyu farmers raise their cattle to gain as large CW as possible while maintaining high meat quality.

## 2.2 Random Forest

Random Forest [1] is a supervised machine learning algorithm proposed by Breiman, which utilizes a collection of decision tree [2] as a weak learner for ensemble learning. Weak learners are less accurate when used alone, and ensemble learning improves the prediction performance by combining those. In the case of Random Forest, a number of decision trees, which are weak learners, are created independently using bootstrap sampling, which creates a dataset for learners by repeating random extraction and restoration. The criterion for dividing each node of the decision tree in a Random Forest consists of two factors: one is the explanatory variable to use, and the other is the threshold on the variable divide the set of samples on the made into two child nodes. The decision tree is created by iterating the division until the end node is created. In the case of regression, each decision tree is given a set of values for explanatory to use for predicting the forget value. The given data set follows the tree by the division criterion of each node to reach one of the leaf node. The regression results for each decision tree are the values for the objective variable of the leaf node it finally reached according to the dividing criterion of the decision tree. The average of those regression results is the regression result in the Random Forest.

There are three characteristics in Random Forest

- (a) Overfitting is unlikely to occur even the number of explanatory variables is large compared to that of samples.
- (b) Compared to deep learning, regression performance is generally less degraded with a smaller data set.
- (c) Possible to calculate the importance of explanatory variables.

Features (a) and (b) allow Random Forest to maintain high estimation accuracy with less deterioration in regression performance, even for data sets with a small number of samples and a large number of explanatory variables. In feature (c), the importance of an explanatory variable effect of variable in predicting values for the variable. In other words, the importance can be considered as an estimate of how much it contributes to the regression effect of the decision tree. In a Random Forest, the closer the calculated importance is to zero, the more irrelevant the variable is to the estimation. Therefore, removing less important variables and training Random Forest again would improve the estimation accuracy. This technique to predict the number of variables with the importance is widely used when using Random Forest.

## 3 The Proposed Method

### 3.1 Data Format

The proposed method suppose the initial body weight and the protein expression profile as explanatory variables. The initial weight is the weight of a calf at when the calf was sold from the breeder to the fattening farmer. In this study, we suppose that the fattening farmer obtain a protein expression profile several time during the fattening period of time. As a result, hundreds or thousands of proteins have their expression levels in the profile. Only CW is applied as the forget trait to be predicted.

Formally, let  $B$  be the set of Wagyu samples. Let  $W_b$  be the initial weight of Wagyu  $b \in B$ . Although the protein expression profiles as a single set of profile, i.e, if we have 6 sets of profiles, each of which includes 100 proteins, we treat it as a single profiles that include 600 proteins. The objective variable is CW and the CW value of  $b \in B$  is written as  $C_b$ . For convenience, the set of initial weights for all Wagyu  $b \in B$  is written as  $W$ , and the set of CW value as  $C$ .

### 3.2 The Idea

This study proposes a method for predicting the carcass weight of cattle using the initial weight and the protein profiles. In this case, the first step would be to apply general methodologies such as the multiple regression analysis or the machine learning. However, these methods balance the nature of each prediction method, and basically do not allow to combine only good characteristics of multiple methods properties. Although some learning methods such as ensemble learning exist that combine multiple machine learning methods, this kind of method takes the average of all prediction results, so the characteristics of the individual methods would not be fully exploited.

On the other hand, it is known that the correlation between CW and initial body weight is large as 0.7 or 0.8. If we can construct a prediction method that takes advantage of this correlation, we would make more accurate prediction than conventional methods. To achieve this, instead of taking ensemble of multiple methods. We adopt a two level learning strategy. We first use simple regression to predict CW to catch up with linear correlation. Then, after that, we apply machine learning to compensate the prediction performance by making use of non-linear information hidden in the data set.

The Random Forest method is chosen as the machine learning method to be applied. In our study, the number of the explanatory variables is considered to be very large. Random forests are less prone to make overfitting while achieving high performance, even when the sample size is small and the number of explanatory variables is large. This characteristic enables us to perform prediction without being bothered by the number of dimensions of the data set.



### 3.3 Wagyu Carcass Weight Prediction Method Combining Regression and Random Forest

The proposed method predicts CW from the initial weight and the protein expression profile given. Our method is divided into a learning phase and a prediction phase. In the learning phase, the model parameters of the regression and Random Forest are determined. The learning phase consists of the following two steps.

**step1** Using the initial weight, we estimate the parameters in the single linear regression analysis.

**step2** We retrieve errors in the single regression from Step 1, and construct a learning model to predict the error from the protein expression profile. For this, Random Forest method is used, i.e., we learn the model parameters of the Random Forest, to predict the amount of errors in Step 1.

In step 1, a single regression analysis is performed using i.e., the initial weight and CW. We set the model equation for a single regression analysis as  $C = \alpha W + \beta$ , then in step 1 the optimal model parameters  $\alpha$  and  $\beta$  are obtained. In step 2, we train a model of Random Forest to predict the "Residual" in step 1. Here, the residual  $r_b \in R$  of the Wagyu  $b$  is expressed as  $r_b = C_b - C'_b$ , where  $C'_b$  is the predicted value of CW for Wagyu cattle  $b$  obtained in step 1. in the model calculated in step 1. The explanatory variable applied in step 2 is the expression profile and the initial weight.

On the other hand, in the prediction phase, each model learned in the learning phase predicts CW with the given data set explanatory variables. This prediction phase consists of the following three steps.

**step1** Calculate the prediction value for CW using the single regression analysis would obtained in step 1 of the learning phase, i.e., for a given initial for the target cattle, we obtain the prediction value for the CW value of the cattle.

**step2** Calculate the predicted residual values by inputting the initial weight, and the protein profiles into the Random Forest learner.

**step3** The final CW prediction is performed by adding the error compensation value obtained from the Random Forest model to the prediction value for CW obtained by the single regression in step 1.

In step 1, the predicted value of Wagyu  $b$ 's CW, i.e.,  $C'_b$ , is calculated as  $C'_b = \alpha W_b + \beta$  with the value of the initial weight  $W_b$  of the cattle sample  $b$ , where model parameters  $\alpha$  and  $\beta$  are obtained in the learning phase.

In step 2, we obtain the prediction value  $r'_b$  for the error called residual from the Random Forest model that compensate the prediction performance of regression. this is done by inputting the initial weight  $W_b$  and the protein expression profile for the target cattle  $b$  into the Random Forest model built in the learning phase. In step 3, we calculate the final prediction value for

CW of cattle  $b$  as  $C'_b + r'_b$ , which is the sum of the prediction values of the two prediction methods executed in step 1 and 2. The above learning and prediction phases predict CW of a target cattle  $b$ .

## 4 Evaluation

### 4.1 Evaluation Method

For evaluation, we used `sklearn.ensemble.RandomForestRegressor` and `sklearn.linear_model.LinearRegression` of `scikit-learn`, an open-source machine learning library in Python. The evaluation method compares the accuracy of the proposed method with existing general prediction methods. The Mean Absolute Error (MAE) is used as the evaluation criterion. Performance comparisons are under variation of the value of  $k$ .

### 4.2 Comparison Methods

The comparison methods are the most basic analytical methods, the multiple regression analysis and Random Forests. When performing a Random Forest, we once performs Random Forest to compute the importance value, and narrowed down the variables by the importance value until the number of variables reached  $k$  and performed the Random Forest again. For multiple regression analysis, we used LASSO [4], a variable selection method in multiple regression analysis, to select the optimal set of variables that consists of  $k$  explanatory variables. We used CW as the objective variable, and initial body weight and a protein expression profiles as the explanatory variables to input the multiple regression analysis. When calculating the MAE with each method, we performed a leave-one-out cross-validation and averaged the values over the number of samples,  $N$ .

### 4.3 Evaluation Results and Consideration

Table 1 shows the lowest MAE for each method, and the number of variables for each lowest MAE. Figure 1 shows the transition in MAE with the various number of variables  $k$  for each of the three methods. In figure 1, the horizontal axis represents the number of variables used for each method, i.e.,  $k$ , and the vertical axis represents the MAE of CW estimated by each method. Figure 1 shows that the proposed method has a smaller MAE than the other methods, even for the same number of explanatory variables. This means that under the same conditions, i.e., data set, and the number of variables, the proposed method is superior to the comparison methods in terms of prediction accuracy.

Table 1: Number of Variables when Recording Minimum MAE

using method	minimam MAE	number of variable
proposed method	21.9617	6
simple Random Forest	25.6906	11
multiple regression analysis	24.3588	8

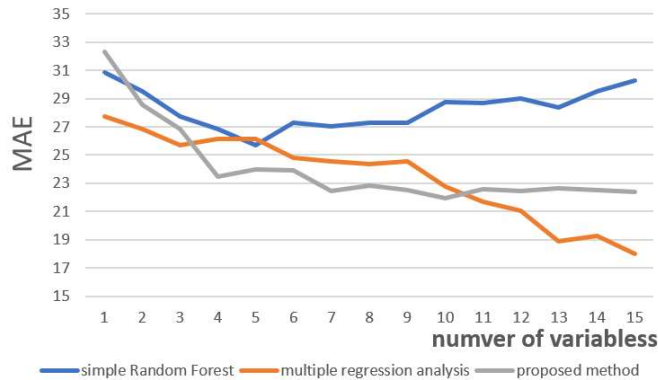


Figure 1: MAE Change of the Three Methods

Figure 1 shows that in the proposed method and simple Random Forest, the MAE gradually decreases until the number of variables reaches that of the smallest MAE. This means that the small number of explanatory variables have the ability to reduce MAE and the effective set of variables can be retrieved when we select explanatory variables with respect to the importance value. Narrowing variables by the importance effectively work in improving the accuracy of MAE estimation. On the other hand, once the optimal MAE has past, MAE rather decreases even though there are as many as 700 explanatory variables. This means that most of the explanatory variables have a low ability to reduce MAE.

The reason why the performance of the proposed method is higher than that of the simple Random Forest owes to the treatment of the initial weight with regression. The initial weight is known to be very highly correlated with CW, and the prediction based on regression analysis captures different characteristics from those based on Random Forest. Figure 1 shows that the importance of initial weight is outstandingly high. In addition, the prediction accuracy of Random Forest is higher when trained with one variable, and the prediction accuracy of the proposed method increases rapidly when the number of variables is increased. This is because the initial weight and the protein profiles capture the common features in Random Forest. On the other hand, the proposed method used the initial body weight in the regression analysis. Therefore, we believe that the proposed method captured characteristics that were not captured by the Random Forest. Table 2 shows the top five variables in the importance and their values in the simple Random Forest.

Table 2 shows that the importance of the initial weight is

Table 2: Importances of Comparison Methods in Simple Random Forests

variable name	importance
initial weight	11.16
protein A	1.644
protein B	1.529
protein C	0.8072
protein D	0.6566

actually greater than the other variables. In the case of the simple Random Forests, the results are averaged because Random Forests is performed on all variables. The comparative method, i.e., the simple Random Forest, process all variables randomly, and so the results are averaged. On the other hand, in the proposed method, the initial weight, which is of the highest importance, is fixed for its effect in regression analysis. In addition, the error, which includes the rest of the effects, is predicted by Random Forests, so that the error is reduced.

## 5 Conclusion

This study proposed the method that combines the regression analysis with Random Forests for Wagyu quality prediction. The performance of the proposed method was evaluated by comparing the prediction accuracy with the existing methods. Among the six major traits of Wagyu cattle, CW was found to be highly correlated with the initial weight. Therefore, we tried to explain the effect of initial body weight on regression analysis and complement the rest by the protein expression profiles.

In the evaluation, the prediction accuracy of the proposed method was investigated by applying real Wagyu data to the proposed method. By ranking variables according to their importance and applying the high-ranked variables to the proposed method, high prediction accuracy was achieved. The results show that the proposed method shows the highest prediction accuracy when the number of variables used is small. There are two issues that need to be addressed for the future, which are (1) the application of the proposed method to other data sets, and the evaluation of prediction accuracy based on the timing of serum collection. As for (1), validation on multiple datasets is needed to show that the current results do not the characteristics specific to the current data set, but the universal characteristics which can be applied to the whole Wagyu cattle. (2) The latter consideration of the timing of collection is necessary to test the practicality of the proposed method. In order to make use of the results of this prediction in practice, a highly accurate forecast is needed. In other words, instead of using all of the protein expression levels at the six time points obtained from the whole period of raising, we should make prediction with the data obtained in the early period of raising.

## Acknowledgment

This work is supported by “Program for Promotion of Stockbreeding” of JRA (Japan Racing Association).

## REFERENCES

- [1] L Breiman: “RandomForst,” Machine Learning, vol.45, pp. 5-32 (2001).
- [2] L. Breiman, J. Friedman, R. Olshen, and C. Stone: “Classification and Regression Trees,” Chapman and Hall/CRC, Published January 1st, pp. 168 (1984).
- [3] LC. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold: “Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a newconcept for consistent and accurate proteome analysis,” Molecular & Cellular Proteomics (2012).
- [4] R. Tibshirani: “Regression Shrinkage and Selection via the Lasso,” Journal of the Royal Statistical Society, Series B (Methodological), Vol. 58, No. 1, pp. 267-288 (1996).

