# International Journal of

# Informatics Society

Informatics Society

**Aims and Scope**

The purpose of this journal is to provide an open forum to publish high quality research papers in the areas of informatics and related fields to promote the exchange of research ideas, experiences and results.

Informatics is the systematic study of Information and the application of research methods to study Information systems and services. It deals primarily with human aspects of information, such as its qu ality and value as a resource. Informatics also referred to as Information science, studies t he structure, algorithms, behavior, and interactions of natural and a rtificial systems that store, process, access and communicate information. It also develops its own conceptual and theoretical foundations and utilizes foundations developed in other fields. The advent of computers, its ubiquity and ease to use has led to th e study of info rmatics that has computational, cognitive and social aspects, including study of the social impact of information technologies.

The characteristic of informatics' context is amalgamation of technologies. For creating an informatics product, it is necessary to integrate many technologies, such as mathematics, linguistics, engineering and other emerging new fields.

# Guest Editor's Message

## Naoya Chujo

Guest Editor of Thirty-seventh Issue of International Journal of Informatics Society

We are delighted to have the Thirty-seventh issue of the International Journal of Informatics Society (IJIS) published. This issue includes selected papers from the Fourteenth International Workshop on Informatics (IWIN2020), which was held online, Sept. 10-11, 2020. The workshop was the fourteenth event for the Informatics Society, and was intended to bring together researchers and practitioners to share and exchange their experiences, discuss challenges and present original ideas in all aspects of informatics and computer networks. In the workshop 24 papers were presented in seven technical sessions. The workshop was successfully finished with precious experiences provided to the participants. It highlighted the latest research results in the area of informatics and its applications that include networking, mobile ubiquitous systems, data analytics, business systems, education systems, design methodology, intelligent systems, groupware and social systems.

Each paper submitted to IWIN2020 was reviewed in terms of technical content, scientific rigor, novelty, originality, and quality of presentation by at least two reviewers. Through those reviews 17 papers were selected for publication candidates of IJIS Journal, and they were further reviewed as a Journal paper. We have three categories of IJIS papers, Regular papers, Industrial papers, and Invited papers, each of which were reviewed from the different points of view. This volume includes five papers among those accepted papers, which have been improved through the workshop discussion and the reviewers' comments.

We publish the journal in print as well as in an electronic form over the Internet. We hope that the issue would be of interest to many researchers as well as engineers and practitioners over the world.

**Naoya Chujo** received his B.E. in applied physics, M.S. in information science and Ph.D. degrees in electrical engineering from Nagoya University in 1980, 1982 and 2004. He joined Toyota Central R&D Labs. in 1982. He has been a professor at the Department of Information Science, Aichi Institute of Technology since 2010. His research interests are in the area of embedded system and automotive electronics. He is a member of IEEE, IPSJ, IEICE, IEEJ, and Informatics Society.

**Regular Paper**

# A Study on Analysis of Viewers' POV and Presentation to Broadcasters 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 problem definition and approach in this study. Section 4 describes the related work and discuss the necessity of this study. Section 5 describes our 360-degree Internet live broadcasting system and its implementation. Section 6 describes the hypotheses about the viewer's viewing behavior and their testing. Section 7 describes an overview of the algorithm and its preliminary experiment. Chapter 8 describes improvements of the algorithm. Section 9 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 10 describes a discussion and Section 11 summarizes this study.

## 2 USECASE

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   PROBLEM DESCRIPTION

In the use case, we envision the use of live 360-degree Internet broadcasting for the broadcast environment. In the 360-degree Internet live broadcasting, the broadcaster is required to provide broadcasting contents equivalent to the conventional ones under the circumstances in which they are hardly able to grasp the state of the viewers. Since the viewers can freely change direction of the view angle, they can get various information from all the directions of 360-degree angles according to their own interests. Simultaneously, the broadcaster has to respond to their various interests. In non-360-degree Internet live broadcasting, the viewers' POV corresponds to the direction of the camera lens and all viewers watch the video at the same angle. The broadcaster can understand what the viewers watch and control the angle of the video arbitrarily by change the direction of the camera. On the other hand, the direction of the omnidirectional camera lens does not show the viewers' POV. This issue has negative effects on communication between the broadcaster and the viewers.

When viewers find something interesting in the broadcaster's surroundings, they make comments to the broadcaster that include pronouns and so on. In linguistics, this behavior is referred to as "deixis" and the response to an instructional expression is expected to be a "Anaphora" that indicates the object corresponding to the instruction [3]. If these confirmations are not properly performed, the correct understanding of the context of the comment from the viewers may be difficult to understand and communication errors may occur. Therefore, we study ways to reduce communication errors based on research on the role and use of non-verbal information in the field of communication.



Figure 1: Equipment of the broadcaster.

## 4   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 [4]. 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 emphasize the need to analyze the viewer's viewing direction to provide a higher quality viewing experience.
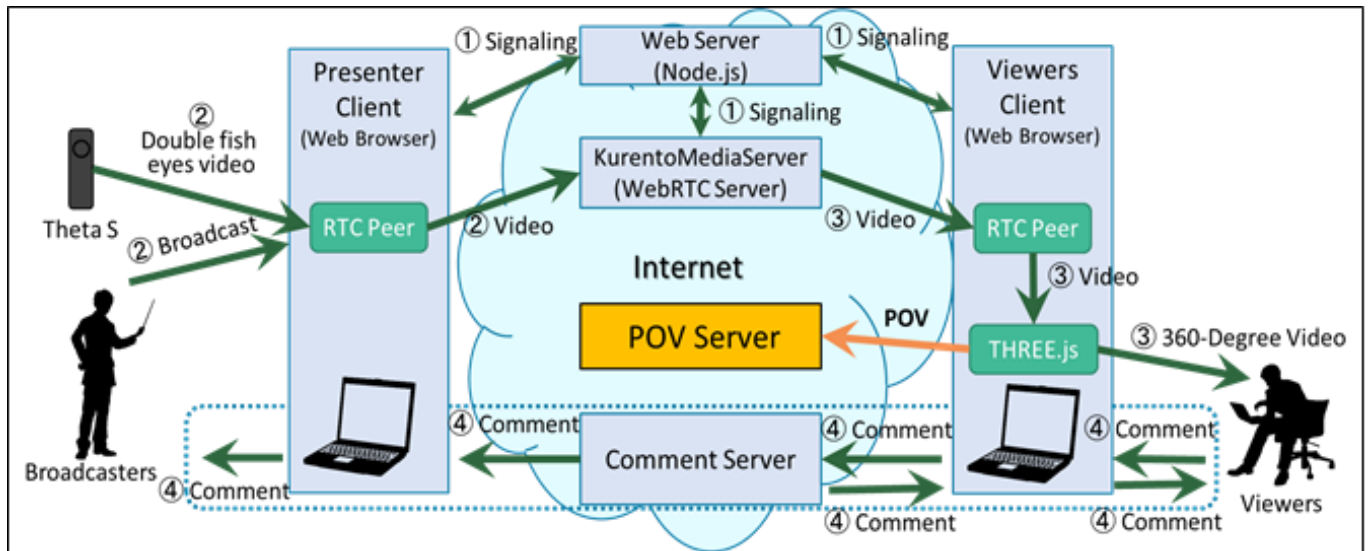
Figure 2: The system architecture of the 360-degree Internet live broadcasting system

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 [5]. 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.

## 5 INTERNET LIVEBROAD CASTING SYSTEM

The experiment environment consists of two components which are a 360-degree Internet live broadcast system and a POV server which collects the viewers' POV information. Figure 2 shows the architecture of the system. We use the Ricoh Theta S as an omnidirectional camera for the system.

The broadcaster accesses a broadcasting system and starts broadcasting by a web browser. Viewers access the broadcasting system and watch the live broadcasting. The viewers' POV information is sent to the POV server periodically.

Various well-known online video distribution services such as YouTube shift from Adobe Flash to HTML5 [6] for the video streaming. We also implemented a 360-degree Internet live broadcast system by adopting WebRTC that realized the video streaming on HTML5. We used the Kurento Media Server [7] for the video streaming server. The WebRTC is an API to provide real-time communication functions such as voice communication and video distribution without requiring any plug-in and installation of special software. We used a javascript library called Three.js [8] for processing images. The image acquired from the omnidirectional camera is mapped to a spherical object by using the library. The viewers' POV are managed by the angular coordinates of the two axes which are acquired every 100ms and the data is sent to the POV server. In this system, the POV is represented by the polar coordinate $(\Phi, \theta, r)$ of the spherical surface. The POV server collects the POV information received from the viewers and saves the collected
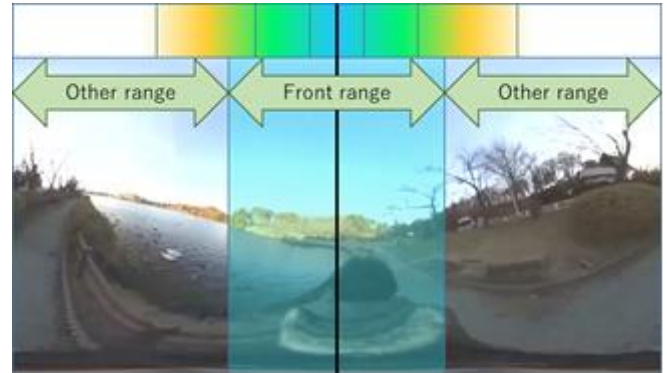


Figure 3: The equirectangular video and the front range.

data in the database. Web server and POV server were implemented using node.js [9].

## 6 HYPOTHESIS TESTING

### 6.1 Hypothesis

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 [5], 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.

## 6.2 Verification Experiment

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. It was found that the viewer's POV was longest directed at a 90-degree range centered on the broadcaster's direction of walk, as explained in Hypothesis 1. 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 [5] 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 3 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.
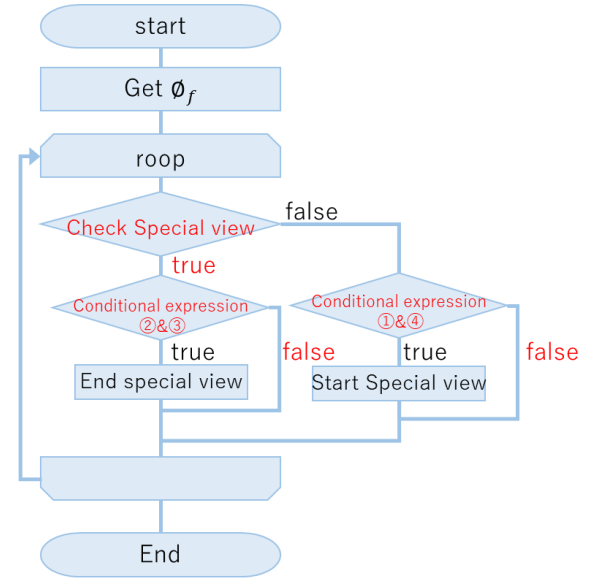


Figure 4: Flowchart of the algorithm



The variable
$$\emptyset_i = \{\emptyset_{i1}, \emptyset_{i2} \dots \emptyset_{in}\}$$
*Horizontal polar coordinates of viewer POV*
forward $= \emptyset_f$
*Horizontal polar coordinates  broadcaster's way*

Conditional expression①
forward $- 45 < \emptyset_{i\,n-9} <$ forward $+ 45$
Conditional expression②
$\emptyset_{i\,n-9} <$ forward $- 45 \,||\,$ forward $+ 45 < \emptyset_{i\,n-9}$
Conditional expression③
forward $- 45 < \emptyset_{i\,n} <$ forward $+ 45$
Conditional expression④
$\emptyset_{i\,n} <$ forward $- 45 \,||\,$ forward $+ 45 < \emptyset_{i\,n}$

Figure 5: The variables and conditional of flowchart.

## 6.3 Algorithm Examination

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. On the other hand, algorithm is not possible detection for other range viewing that less than 1 second, because it becomes possible to detect it 1 second after start of other range viewing. However, data analysis confirms that other range viewing takes place for an average of 11.41 seconds It is believed that it is possible to detect a sufficient. 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 order to implement the algorithm, you need to acquire and compensate the direction of the broadcaster's way using acceleration sensors.

# 7  ALGORITHM

## 7.1  Overview

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 4 shows a flowchart of the algorithm we created. Figure 5 shows the variables and conditional expressions used in the flowchart.

## 7.2  Verification

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.

Table 1: Number and target of broadcasting.

| Broadcast | A Target: Yes Topicality: Yes | B Target: Yes Topicality: No | C Target: No |
|---|---|---|---|
| 1st broadcast | 35 data | 29 data | 9 data |
| | A rowing boat A small shrine near the pond A characteristic building | Pedestrians A flock of birds Swans on the opposite of the pond | Looking around Test operations |
| 2nd broadcast | 41 data | 44 data | 10 data |
| | A wild duck landing on the pond A vinyl house A person shooting a photograph of the pond | Trees by the roadside Bicycles parking around the pond Sunset | Looking around Test operations |
| Percentage | 45.24%  (76/168) | 43.45%  (73/168) | 11.31%  (19/168) |

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. Table 1 shows the number of the three categories and their targets for each broadcast. 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.

## 7.3  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 ±62.5

degrees or more. We decided to redefine the boundaries of the algorithm to ±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 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.

## 8　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 6 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.

### 8.1　Evaluation Experiment

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
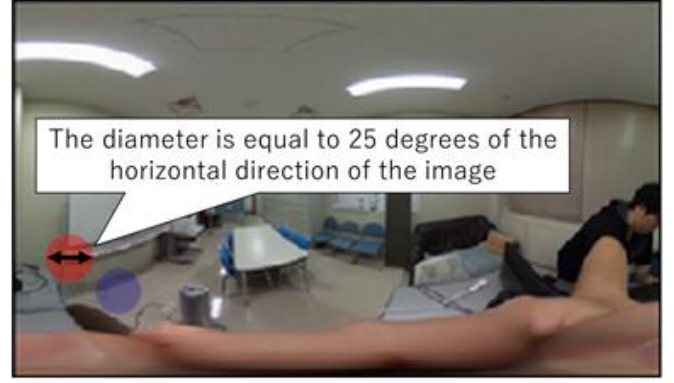


Figure 6: Superimposed POV indication

Table 2: The percentage results of the topical and notice.

| broadcast | Topical | Notice | Persentage |
|---|---|---|---|
| 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%) | |

Table 3: The Results of target objects and reason

| Broadcast | Topical | Target × Number | Reason |
|---|---|---|---|
| 1st broadcast (35) | Yes (20) | Birds × 9 | They were Interesting and can expand the topic. |
| | | Rowing boats × 4 | I turned out viewers' interests in boats. |
| | | Edge of pond × 3 | The shape was easy to expand the topic. |
| | | Other × 4 | I could expand the topic of dog walking. |
| | No (15) | Street tree × 5 | I thought it couldn't expand the topic |
| | | State of pond × 5 | I didn't see anything in the pond. |
| | | Buildings × 4 | It's a just landscape without features. |
| | | Backward direction × 1 | There was nothing in particular behind me. |
| 2nd broadcast (46) | Yes (39) | Frozen pond × 11 | The viewers were interested in it. |
| | | Birds × 9 | They were Interesting and can expand the topic. |
| | | Persons × 9 | A parent and a child enjoyed feeding birds and I thought it had topicality. |
| | | Other × 10 | I could talked about a policeman who asked me questions, persons who took a walk and visited the pond. |
| | No (7) | Frozen pond × 2 | Birds were landing when talking about the birds. |
| | | Cars × 2 | It's a just landscape without features. |
| | | Building × 1 | It's a just landscape without features. |
| | | Other × 2 | I felt it was a meaningless viewing. |

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.

## 8.2    Evaluation Results

From the experimental results, the detecting other range viewing by the improved algorithm were Group A and Group B were 35 and 46. Table 2 shows the results of the interviews with the broadcasters after the experiment if there were the topicality and the reasons for the responses. 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. Among those noticed in the POV presentations, 14 were judged to be topical, and 22 were judged to be topical.

Table 3 shows the results of target objects of viewing and reason. In some cases of other range viewing, the viewers watched same objects at the same time. The ones shown in red are the representative targets of the viewing which were judged to be topical. 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 eye, such as passersby, were mentioned.

The broadcaster explained that these reason for this decision was that he was able to grasp the interest in wild birds and that he was able to get a lot of comments on the topic of police patrols where the POV was shown. Regarding the presentation of the POV which was not noticed during the broadcast, we also received comments that "if I had noticed the interest in the boats, I would have moved a little closer to the boats" and "if I had been able to notice it during the broadcast, I would have used it as a topic of discussion". Even for those which were judged not to be topical, the broadcasters commented that it did not require a big effort to check and did not bother them too much when presented, and that it was helpful to be able to notice the car traffic, even though it was not topical. Therefore, it is expected 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.

## 9    DISCUSSION

### 9.1    Effect of POV Presentation

The reasons for the presence or absence of topicality can be divided into three categories. The first reason is that the broadcaster feels topical that the state of the object or the change in condition of the object. The broadcaster can get a topic from the changing condition of the object. If the state of the object doesn't change, it is expected that it will be difficult to keep talking about it. In the experiment, the broadcaster talked 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.

### 9.2    Future Issues

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.

False positives from the analysis algorithm were classified and methods for improvement were discussed. The items which were detected incorrectly this time can be classified into three patterns. The first pattern was viewing near the boundary line and the POV left from the front range for a very short time. The countermeasure to the first pattern was to extend the front range. We think another countermeasure is to weight the data based on the distance from the bounda-

ry line. The second pattern is stopped the POV control and left in viewing of other range. The countermeasure to the second pattern is to filter by the amount of POV movement. If possible detection of interruptions in operation based on the viewer's POV change while viewing other ranges, it is possible to exclude them from the detection. The third pattern is an unexpected behavior such as checking operations. A possible countermeasure to the third pattern is to filter the POV operation direction. By converting the data, we can analyze the degree of meandering, and this would allow us to determine whether the object of viewing exists.

## 10 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.

## REFERENCES

[1] R. 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] D. M. Grayson, A. 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] S. C. Lee, "Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse", Massachusetts Institute of Technology201 Vassar Street, W59-200 Cambridge, MA, United States.

[4] Y.C. Lin, Y. J. Chang, H. N. Hu, H. T. Cheng, C. W. Huang, M. 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).

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

[6] YouTube Engineering and Developers Blog: YouTube now defaults to HTML5 <video>, https://youtube-eng.googleblog.com/2015/01/youtube-now-defaults-to-html5_27.html, November 7, 2018.

[7] Kurento, http://www.kurento.org/, November 7, 2018.

[8] three.js - Javascript 3D library, https://threejs.org/, November 7, 2018.

[9] Node.js, https://nodejs.org/ja/, November 7, 2018.

**Masaya Takada.** received his master degree from Iwate Prefectural University, Japan, in 2018. He is currently studying software engineering at the post-doctoral program at the University. His research interests include 360-degree Internet live broadcasting. He is a member of IPSJ and IEEE.

**Yoshia Saito**. received his Ph.D. degree from Shizuoka University, Japan, in 2006. He had been an expert researcher of National Institute of Information and Communications Technology (NICT) from 2004 to 2007, Yokosuka, Japan. He was a lecturer from 2007 to 2011 at Iwate Prefectural University and he is currently an associate professor at the University. His research interests include computer networks and Internet broadcasting. He is a member of IPSJ, IEEE, and ACM.

**Regular Paper**

# A Study of Increasing Communication Reliability of Low-cost Field Servers

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. Therefore, it meets farmers' expectation of utilizing reasonable FSs.

*Keywords*: LoRa, Fail-safe, Multi-hop, Ad hoc transmission algorithm, Time synchronization

## 1 INTRODUCTION

The agricultural working population has decreased by approximately 925,000 from 2010 to 2019 in Japan [1]. The situation in the field of agriculture has been dire. Hence, development of Internet of things (IoT) to support rice farming is desired. In recent years, field servers for rice cultivation have been proposed, but they are expensive, and many of them are fixed type that cannot be easily moved, such as holding a large solar panel [2].

We develop a rice cultivation management system using the field servers (FSs) to reduce the workload of farmers [3, 4, 5, 6]. 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. It is known that the higher the position of the FS antenna has the better the propagation characteristics. When the time of harvesting rice, the rice grows to a height of about 1.2 m [7]. On the other hand, considering the ease of movement, the FS should be as compact as possible. Therefore, we set the height of FS to about 1 m. The system has a reasonable price that is less than ten thousand yen per one FS; therefore, rice farmers can make a profit even if they introduce one FS for each rice field [3].
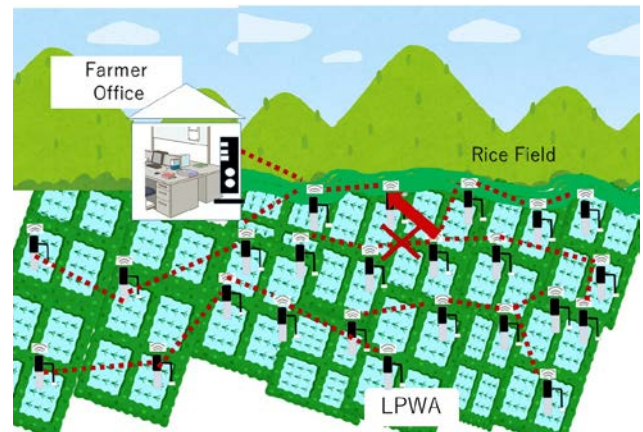
Figure 1: Transmission failure and transmission route change.

The proposed system allows time errors to achieve the reasonable price. We set one step period as tens of seconds to allow for time error. We do not perform precise time management. 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 [4, 5]. Using this algorithm, it is possible to collect all the data in a few minutes, thus achieving a user's request within 5 min. It also runs on batteries for about 6 months, which is the period from rice planting to harvesting.

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.

Dynamic routing in wireless multi-hop networks has been proposed to improve communication reliability [8]. The node corresponding to FS periodically broadcasts a Hello message, exchanges quality information with neighboring nodes, and dynamically determines a route. However, there is a problem that the maintenance cost of the network is high. And it is difficult to use in FS that requires intermittent operation.

Recently, the simultaneous-transmission-type flooding algorithm has been proposed as a method of transmitting data without scheduling [9]. This method can build a stable

and efficient sensor network by repeated flooding without the use of routing. No scheduling is required, so there is no need for updating the network status sequentially. However, this method requires high-precision time synchronization and the apparatus becomes expensive. In Japan, a low cost FS is desired. In addition, when this algorithm is used for FS data collection, 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.

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. The field server for rice cultivation needs to operate outdoors for about 6 months using only batteries. In order to move out of the way of agricultural machinery, the weight of FS must be below about 2 kg. Farmers hate wild birds that are perched on the field servers, so the height of FS needs to be below about 1m.

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 [3]. 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.
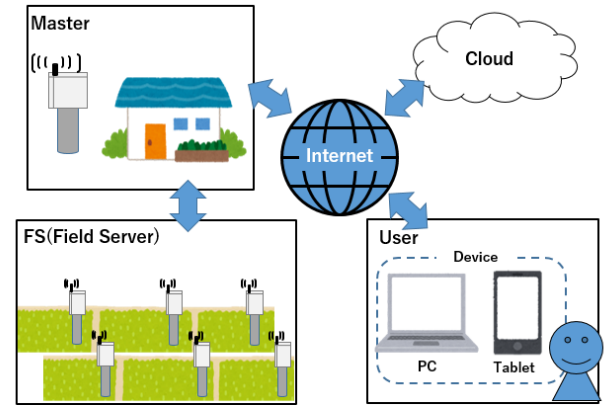


Figure 2: Rice cultivation management system.

In the proposed rice cultivation management system, the FSs are placed in a position where data can be transmitted to the master unit in one hop. Both the master unit and FSs have one LoRa antenna and do not have the function of transmitting and receiving using multiple channels at the same time. In order to realize low cost, FS does not hold expensive parts such as crystals and GPS, and has a time error. The FS must send a water level data to the master unit once an hour. The process which is once an hour of is called one round. It have to be within about 5 minutes to collect all FS data.

In the proposed algorithm, first, the FSs data is collected in the master unit system by using the data collection algorithm. Next, the master unit system confirms the data that has not been received and sends a retransmission request to the FSs. Upon receiving the retransmission request, the FSs send the data to the master unit system using the simultaneous-transmission-type flooding algorithm.

## 3  DATA COLLECTION ALGORITHM

The data collection algorithm is explained [4, 5]. 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. Repeat step4 and step5 until all FSs transmit data and their power has been switched off.
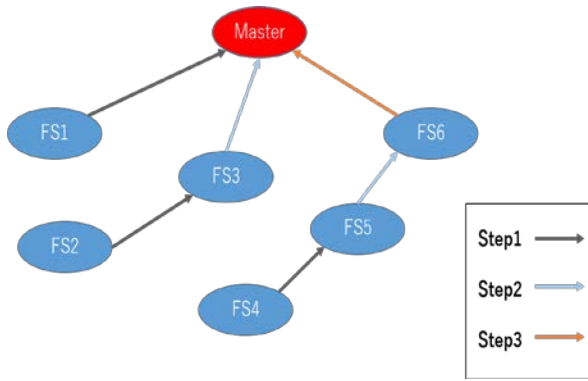
Figure 3: Example of transmission to the master unit system.

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.

Data collection algorithm is made up of two of the processing of the scheduling phase and data collection phase. The detailed scheduling phase is shown below.

```
//Scheduling phase

 1: node={all FSs}
 2:
 3: //find(key)=Find a FS that is closest to key from list.
 4: //findf(key)=Find a FS that is farthest to key from list.
 5: //remove_list(key1)= Remove key from list
 6: //remove_node(key1)= Remove key from node
 7: //add(key1,key2)= Add a transmission schedule from
 8:          key1 to key2 to Scheduling table[period].list
 9:
10:    period=0
11: while (node != NULL) {
12: list = node
12: period++
14: SFS=find(master unit)
15: remove_list(RFS)
16: remove_node(RFS)
17: add(SFS, master unit)
18:        While (list != one or less){
19:            RFS=findf(master unit)
20:            remove_list(RFS)
21:            SFS=find(RFS)
22:            add(RFS, SFS)
23:            remove_list(SFS)
24:            remove_node(RFS)
25:        }
26:  }
```

The detailed data collection phase is shown below. Data collection is done by scheduled. The processing of the master unit and the processing of each FS are showed. The period number of master unit system is globally managed and updated within the master unit system. Similarly, the period number of the FS is globally managed and updated within the FS.

```
// Data collection phase
//Master unit system

1:   while (receive data) {
2:        Save data to database.
3:        }
```

```
// Data collection phase
//FS

 1: // Variable
 2: period number   //The current number of periods
 3:                 // It is calculated from timer.
 4: Scheduling table[].mode   //Scheduling table
 5:                     //(mode: receive, send, sleep)
 6:
 7: data list = sense data  // Sensor data of FS
 8: While(Scheduling table[period number].mod!=NULL){
 9:   if ( Period ! = period number){
10:       State = Scheduling table[period number].mode
11:       Period = Period number
12:       }
13:   if ( State == receive){ // Receive mode
14:       Receive data
15:       Add data to data list
16:       State = NULL
17:       }
18:  else if ( State == send){ // Send mode
19:     Send data list
20:     State = NULL
21:     }
22:     else if( State == sleep){ //Sleep mode
23:     Sleep one period
24:     State=NULL
25:       }
26:  }
```

## 4    SIMULTANEOUS-TRANSMISSION-TYPE FLOODING ALGORITHM

The study of the simultaneous-transmission-type flooding algorithm for the rice cultivation management system is explained in [6]. 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.
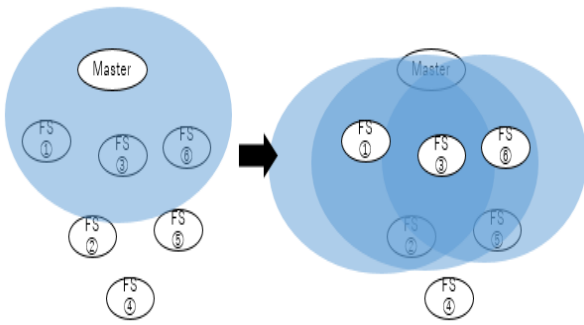
Figure 4: Hierarchy broadcasting from the master unit system.

The detailed simultaneous-transmission-type flooding algorithm from the master unit system to FSs is shown below. The processing of the master unit and the processing of each FS are showed. The master unit system broadcasts the data to all FSs. On the other hand, when each FS receives the data, it broadcasts the data and turns it off. Processing is done asynchronously.

```
// Simultaneous-transmission-type flooding algorithm
//Master unit system

1:   Broadcast data
```

```
// Simultaneous-transmission-type flooding algorithm
//FS

1:  while (receive data) {
2:       Broadcast data.
3:       Power off.
4:       }
```

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. We chose this method because we need to ensure that the transmit request is sent to the untransmitted node and that the data from the untransmitted node is sent to the master unit system. This method is likely to have multiple transmission paths to an FS with a communication failure. In addition, it is a method with a high possibility of receiving data from a node with a communication failure via multiple routes. Therefore, the reliability of communication is high.

## 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 [10, 11]. 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.
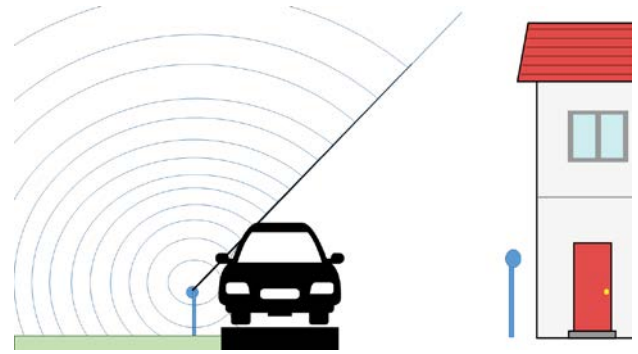


Figure 5: State of radio wave propagation.

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

//Schedule mode
1: Collect data from all FSs using the data
collection algorithm
//Fail Safe ad hoc mode
2: While(){
3:   if (Was the master unit system able to collect data
        from  all FSs?)
     // Broadcast EndMessage
4:      then{
5:        Master unit system transmit the retransmission
         scheduling information by using the simultaneous-
         transmission-type flooding algorithm.
```

```
 6:            Break;}
         // Broadcast NoconnectedList
 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.
                 }
```



Figure 6: Microcomputer time error.

## 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 [12].

The proposed method is a method that allows time errors. Therefore, detailed verification is required to check the operation. Since the FS we created consumes a current of about 50mA when transmitting data, we can verify the system operation by checking the current waveform. In order to measure the current waveform, it is necessary to give an accurate time to the battery-powered measuring instrument for observation. Therefore, one battery-powered measuring instrument was installed in every FS, and the time was given to the battery-powered measuring instrument via LAN. This experiment was conducted indoors. Figure 7 shows the experimental environment. Since the master unit and the battery-powered measuring instrument are connected to the same LAN, the master unit and the battery-powered measuring instrument can be completely synchronized. Therefore, accurate measurement is possible.

An experiment of transmission delay was also conducted depending on the transmission distance. This experiment was conducted outdoors in Nonoichi City, Ishikawa Prefecture.
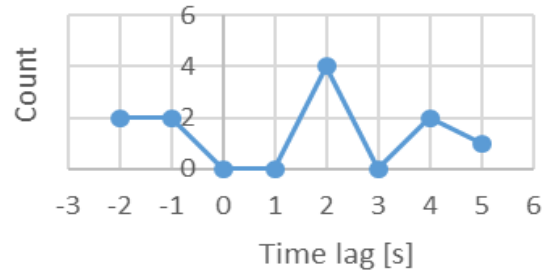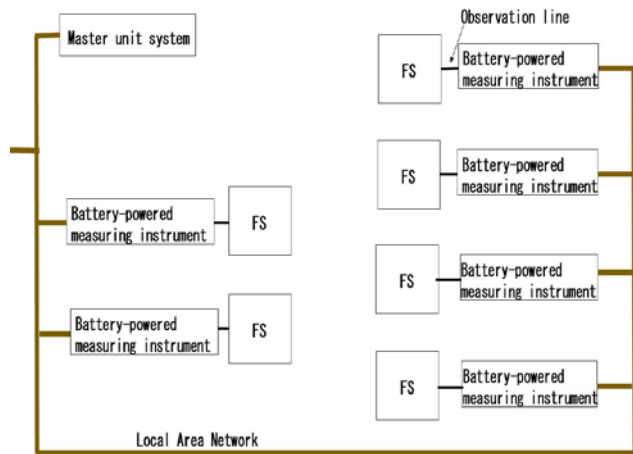


Figure 7: Experimental environment.

### 6.2 Experimental Results of the Data Collection Algorithm

The experimental results of the data collection algorithm are shown in Figure 8. We explain the results using the example of Figure 3 which has six FSs. Figure 8(a) shows the sequence diagram of the example. In addition, Figure 8(b) shows the current waveform of FSs 4, 5 and 6. Look at the circled area of Figure 8(b). 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. Our device has a time error, so it cannot start at the same time. It can be confirmed that the transmission interval is not constant and is asynchronous. 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.

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 master unit system 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.

Even with the same scheduling, the current waveform will be different each time. This is because the variation of the device is different each time. In order to execute the operation correctly, it is important to determine the step period so that there is no overlap in each step, including variations.
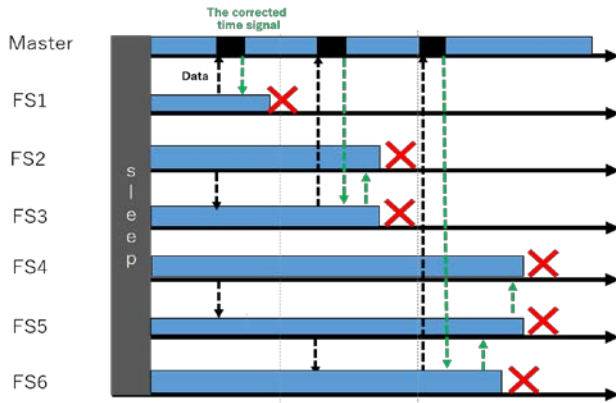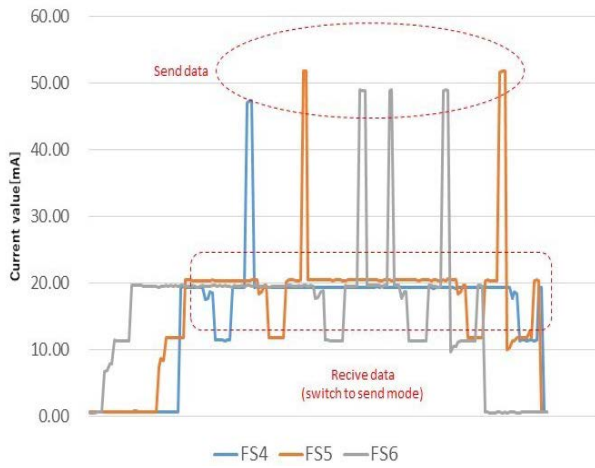
Figure 8(a): Sequence diagram of Figure 3.



Figure 8(b): Current measurement result of Figure 3.



Figure 9(a): Sequence diagram of the simultaneous-transmission-type flooding algorithm.
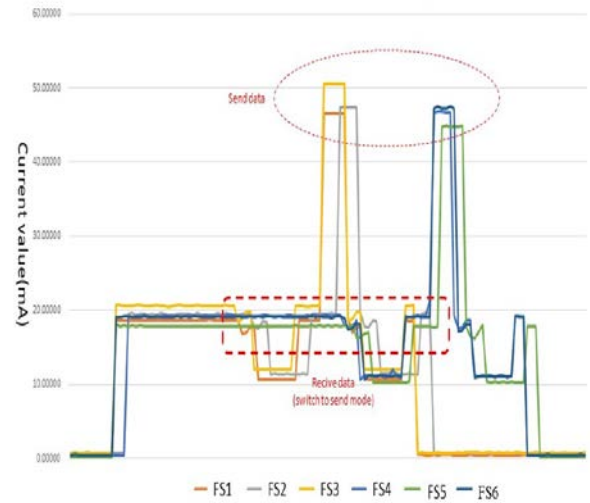


Figure 9(b): Current waveforms of six FSs example.

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

The experimental results of the simultaneous-transmission-type flooding algorithm are shown in Figure 9. Figure 9(a) shows a sequence diagram. First, 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 about 10 s. Second, since there are individual differences time error 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 and turn off the power. Figure 9(b) shows current waveforms of a six FSs example. It can be confirmed that the transmission interval is not constant and is asynchronous.
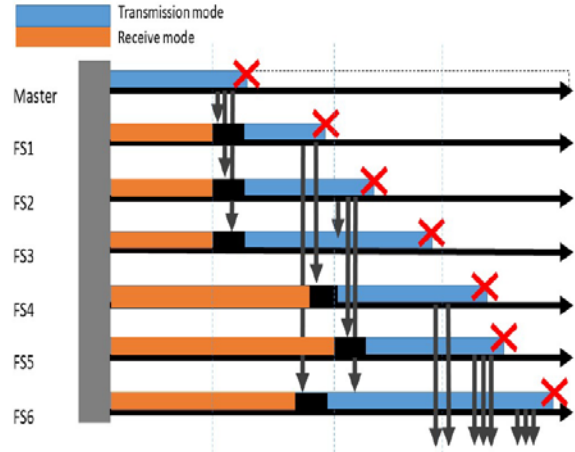
Figure 10 shows the measurement results of transmission from random points in Nonoichi City to the gateway on the roof of the Kanazawa Institute of Technology Library Center. The horizontal axis represents the RSSI. The vertical axis represents the transmission delay. Nonoichi City is located in the center of Ishikawa Prefecture and is a completely flat land with no mountains or sea. In addition, it is 4.5 kilometers east-west and 6.7 kilometers north-south, which is a size that can cover the city with LPWA. As you can see from the Figure 10, the transmission delay also varies. And there is a delay of a few seconds. If you run the simultaneous-transmission-type flooding algorithm under these conditions, the process may not be completed within the step period if several multi-hops occur. Since the system is configured within the range where it can be directly transmitted to the gateway, it is unlikely that more than one multi-hop will occur. In addition, the step period is determined so that there is sufficient time to spare even if two multi-hops occur, so the process is designed so that it

will be completed in time. However, exceptionally, if the processing is not completed within the step period, it gives up and shifts to the phase in which the master unit confirms the unreceived data, which is the next treatment.

The proposed method assumes that all FSs can communicate without multi-hop. Therefore, even if there is some problem and the transmission from the master unit system cannot be received, the transmission from the plurality of FSs that received the transmission from the master unit system can be received. Since there are multiple transmission routes, the probability of not receiving data in two multihops is very low. Our method is a reliable method. Therefore, the step is set to about twice the worst transmission delay. If communication is not possible by this highly reliable proposed method with multiple communication paths, it is considered that there was a failure peculiar to that time, such as the field server being pulled out for work, and it is decided to proceed to the next process.
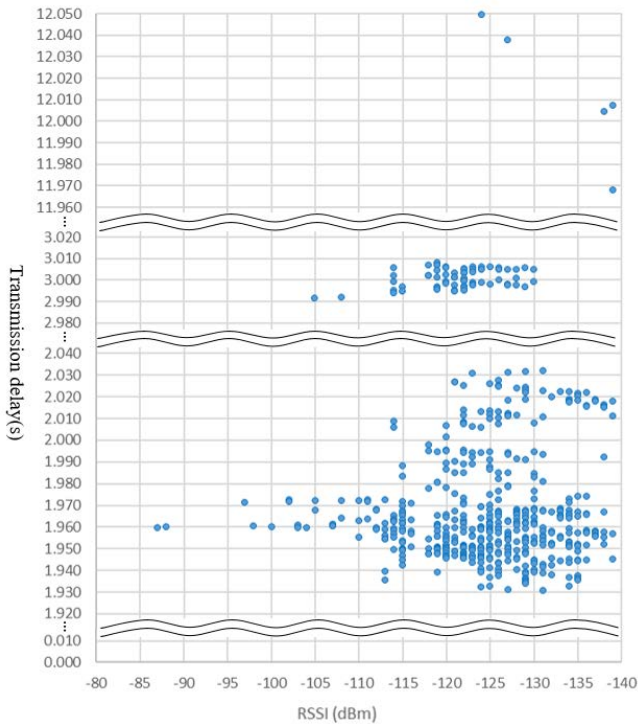


Figure 10: Transmission delay variation.

## 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 made an environment where data transmission errors occur intentionally and conducted experiments. Two examples are shown to display the working of the algorithm.

|  | #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 |

Table 1: FS channel settings
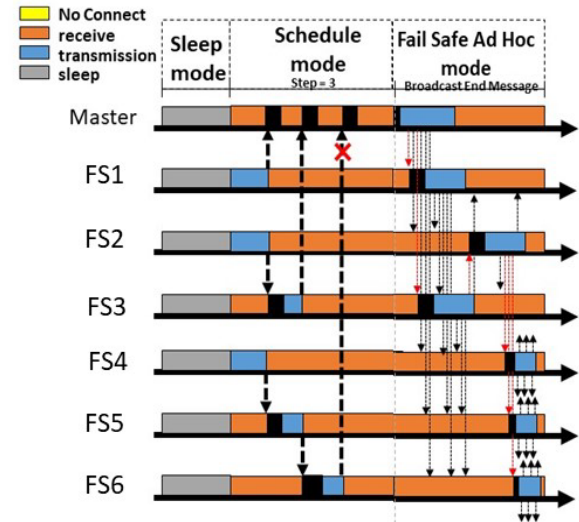


Figure 11(a): Sequence diagram where retransmission processing does not occur.
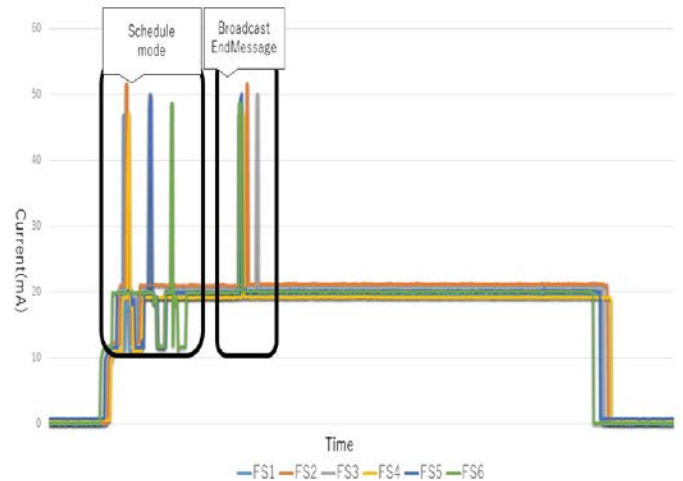


Figure 11(b): Current waveforms of a Figure 11(a).

First, the experimental results of Figure 3 are shown when the operation is normal without retransmission. Figure 11 (b) is a sequence diagram, and Figure 11 (c) is an example of the corresponding current waveform. The data collection algorithm was executed, unreceived data was confirmed in the master unit system, and since there was no unreceived data, the completion information was sent by the simultaneous-transmission-type flooding algorithm.

Figure 12(a): Test example 1.



Figure 12(b): Sequence diagram of Figure 12(a).

Next, the experimental results of resending are shown. An experimental result of an example where transmission from FS 6 to the master unit system is disabled is shown in Figure 12. Figure 12(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 an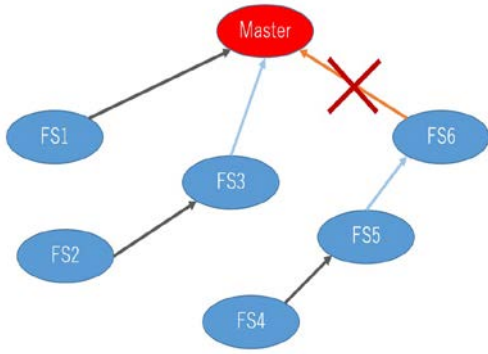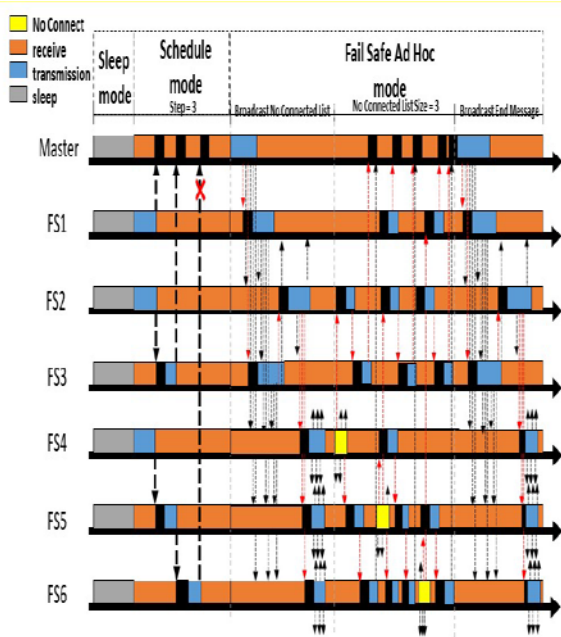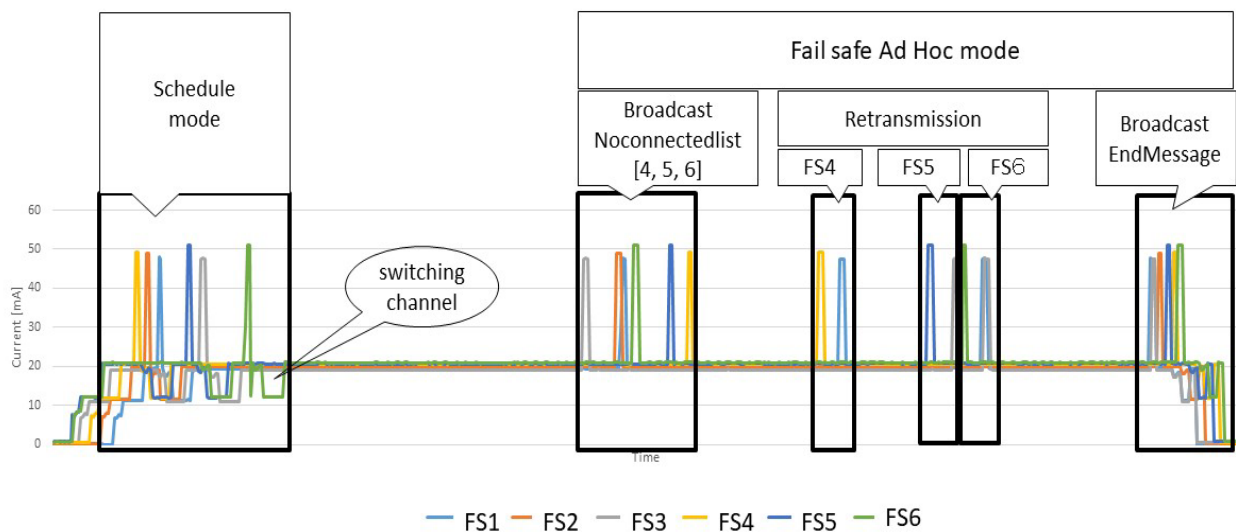d FS 5 sends data to FS 6. In step 3, FS 6 sends data to the master unit system. However, communication is not possible due to the large distance between FS 6 and the master unit system.

The master unit system determines that the communication is not complete, and broadcasts the retransmission scheduling information of FS 6, FS5 and FS4 for which communication is not completed, using the simultaneous-transmission-type flooding algorithm. In the next step, FS 6, FS5 and FS4 retransmits data using the simultaneous-transmission-type flooding algorithm. After that, 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 12(b) shows the sequence diagram of Figure 12(a).

Figure 12(c) shows the measurement results of the current waveform during operation in Figure 12(b). The current at the time of transmission is approximately 50 mA. We compare the position of the transmission on Figure 12(c) to the blue box position showing transmission in Figure 12(b). Thus, we can confirm that the transmission position is the same. It is a mechanism to erase the data after transmission due to the problem of the memory amount of the FS CPU. Therefore, retransmission processing is performed with FS6, FS5, and FS4. In the case of this example, if the problem of memory amount is solved, it is possible to send the data of FS6, FS5, and FS4 only by resending FS6. Memory problems can easily be resolved when we can use an expensive CPU. However, the system we propose cannot use an expensive CPU. Therefore, it is a necessary process to realize a low price FS.
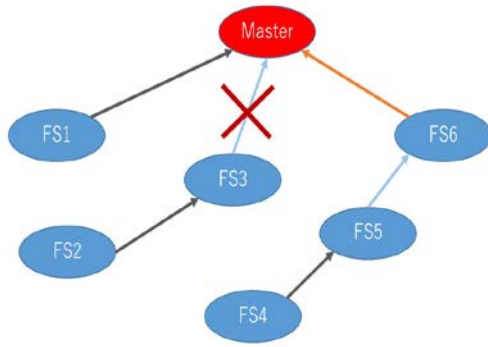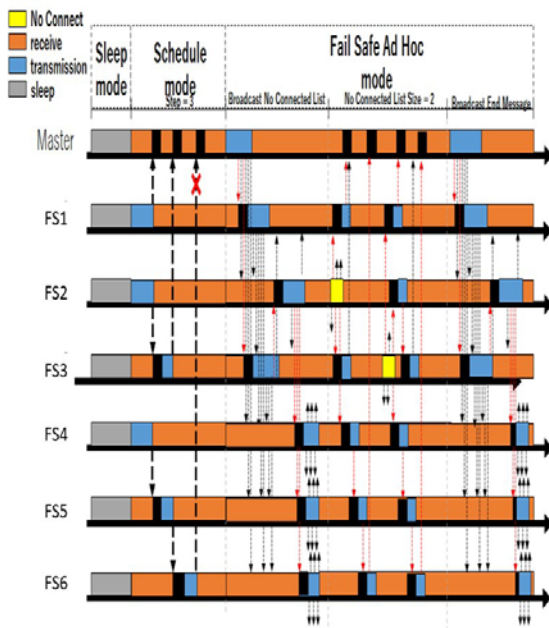


Figure 12(c): Current waveform of Figure12 (a).
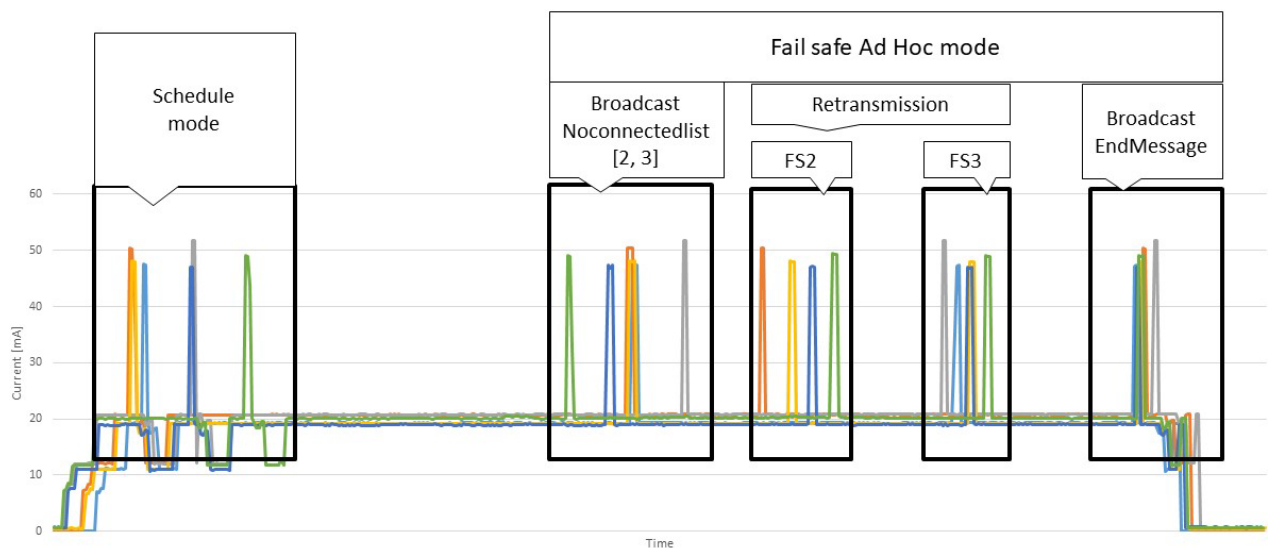
Figure 13(a): Test example 2.



Figure 13(b): Sequence diagram of Figure 13(a).

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 13(a). Figure 13 (b) is a sequence diagram, and Figure 13 (c) is an example of the corresponding current waveform. 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.

After that, 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 according to scheduling. 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 [4]. 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.



Figure 13(c): Current waveform of Figure 13(a).

To achieve low cost, FS does not hold expensive parts such as crystals and GPS, and has a time error because of device variations. The amount of CPU memory is also minimal. The proposed algorithm allowed device time variations and transmission delay variations. Also, each variation is several seconds. Therefore, it is necessary to set the time for one step so that this variation can be tolerated. In the experiment shown in this paper, one step was set to about 1 minute for the purpose of clearly showing each process. In other words, if their variation is small, the time for one cycle can be shortened and the processing time can be shortened. And many FSs can be connected to one master unit system. Our goal is to make a low-priced FS, which meets the purpose of collecting the data requested by the user in about 5 minutes under the conditions that there are device restrictions to reduce costs. We think that the purpose is satisfied.

# 7  CONCLUSION

We develop a rice cultivation management system using the field servers (FSs) to reduce the workload of farmers. Price reduction is essential for the introduction of FSs. 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 proposed method is to take advantage of the fact that the data collection algorithm has few untransmitted data, and to adopt the simultaneous-transmission-type flooding algorithm, which is a robust algorithm with an increased number of transmissions. The conventional simultaneous-transmission-type flooding algorithm requires highly accurate time synchronization; however, by avoiding this perfect simultaneous transmission, it was possible to successfully operate with a low cost FS.

As a result of the experiment, we have confirmed that the master unit system was recieved all data of FSs. In addition, the current waveform is shown to show the operation accurately. This protocol is robust for the rice cultivation management systems, because the unreceived node is retransmitted using multiple routes. Therefore, it meets farmers' expectation to utilize a reasonable FS.

# REFERENCES

[1] Ministry of Agriculture, Forestry and Fisheries, https://www.maff.go.jp/j/tokei/sihyo/data/08.html

[2] "IIJ's Efforts to Promote LoRaWAN? in Agricultural IoT," https://www.iij.ad.jp/en/dev/iir/pdf/iir_vol47_focus1_EN.pdf (Access: 2021/1/5).

[3] 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).

[4] Mikiko Sode Tanaka, "Proposal of IoT Communication Method for the Rice Field,"

International Journal of Informatics Society (IJIS) (2020)

[5] 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.

[6] 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.

[7] Gao, Zhenran; Li, Weijing; Zhu, Yan; Tian, Yongchao; Pang, Fangrong; Cao, Weixing; Ni, Jun. 2018. "Wireless Channel Propagation Characteristics and Modeling Research in Rice Field Sensor Networks." Sensors 18, no. 9: 3116.

[8] Yukio Okada, Kazuo Dobashi, Naoki Umeda, Takayuki Sasaki, "Scalable Multi-Hop Network System for Smart Meters," Panasonic Technical Journal, Vol. 57, No. 4, Jan. 2020.

[9] Federico Ferrari, Marco Zimmerling, Lothar Thiele, and Olga 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.

[10] Takaki Yumi, Minami Hiroaki, Ohta Chikara, Tamaki Hisashi, "A Study on Penetration Strategy of On-board Unit Considering the Type of Vehicles and the Obstacle Shadowing in V2V Communications," Transaction of the Japan Society for Simulation Technology, Vol.7, No.4, pp.125-133, 2015.

[11] Mitsubishi Electric System & Service Co., Ltd., "920MHz band wireless unit SWL90 series antenna laying manual." https://www.melsc.co.jp/business/wireless/swl90/pdf/x903130602c.pd (Access: 2020/11/04).

[12] Yumeto Kojima, Mikiko Sode Tanaka, "Current measuring instrument for field management field server using LoRa," 2018 IEICE Society Conference, 2018/9/11-14.

**Mikiko Sode Tanaka** received Dr. Eng. degrees from Waseda University in Fundamental Science and Engineering. She joined NEC Corporation, NEC Electronics Corporation, and Renesas Electronics Corporation. She is Associate Professor of International College of Technology, Kanazawa. Her research interests include wireless communications, AI chip, and personal authentication. She is a member of IPSJ (Information Processing Society of Japan), IEICE (Institute of Electronics, Information and Communication Engineers) and IEEE (Institute of Electrical and Electronics Engineers).
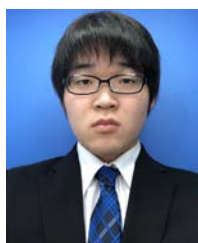


**Yuki Okumura** received a Bachelor's degree from Kanazawa Institute of Technology and graduated in 2021. He joined Hokutsu Co., Ltd. He is currently engaged in construction management.
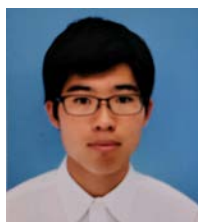


**Sota Tatsumi** received a Bachelor's degree from Kanazawa Institute of Technology and graduated in 2021. He joined OPTiM Co., Ltd. He is currently engaged in CIOS development.



**Shogo Ishii** received a Bachelor's degree from Kanazawa Institute of Technology and graduated in 2021. He joined Hokuriku Telecommunication Network Co., Ltd. He is currently engaged in the construction of telecommunications equipment.



**Tatsuya Kochi** received a Bachelor's degree from Kanazawa Institute of Technology and graduated in 2021. He joined West Japan Railway Company. He is currently engaged in the management of electrical equipment.

# Estimating a Specific Position Related to an Event for Deflation Detection

Takuma Toyoshima[*], Takuo Kikuchi[**], Masaki Endo[**], Shigeyoshi Ohno[**] and Hiroshi Ishikawa[***]

[*]ICT Production Support, Polytechnic Center Shiga, Japan
[**]Division of Core Manufacturing, Polytechnic University, Japan
{kikuchi, endou, ohno}@uitec.ac.jp
[***]Graduate School of System Design, Tokyo Metropolitan University, Japan
ishikawa-hiroshi@tmu.ac.jp

*Abstract* - Sensors are useful to track a person's movement at an event. Physical sensors are expensive to install and maintain. Therefore, instead of physical sensors, social network service (SNS) users can be treated as sensors to observe real-world events. To do so, many data are required. However, for information protection, few SNSs retain accurate location information. To ascertain the movement of a person at an event using a social sensor, which costs less than a physical sensor, one must infer the location information of the social sensor. Therefore, we assess a method of estimating position information related to a specific event. Estimation accuracy was evaluated using actual data from tweet messages recorded in Chiyoda ward, Tokyo. For SNSs with accurate location information, cross-validation is used to evaluate the location estimation method. However, because few SNSs have accurate location information, we also use SNSs with location information for each city, ward, town, and village. In this case, the position estimation result is considered by evaluating the accuracy of the SNS as a sensor. By estimating the position, one can increase the SNS data posted near a specific event. Furthermore, using this SNS data, one can track the movement of people at an event more accurately.

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

## 1 INTRODUCTION

Computerized devices and systems used in modern society provide many means for real-time acquisition of diverse information. One means is Twitter, a microblogging service that shares short sentences called "Tweets" of 140 or fewer characters. The service is widely used throughout the world, as it is in Japan. Many users regard it as a medium by which they can post information continually and casually. Posting of location information can be done easily via a smartphone with so-called geotags. This social medium can therefore 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 expensive physical sensors.

Social sensors can elucidate a situation in real time even if one is not present on the scene. For example, if one can estimate the best time to view cherry blossoms and autumn leaves before visiting a place, one could actually go to a place with no concern that cherry blossoms have not yet bloomed or that they have already fallen. If public transportation is halted and a person knows that people are crowding into stations, then that person can avoid those crowds. If a person knows that congestion in an area has eased, then the person could plan a route through 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 settles down, it might not be so interesting. Alternatively, to avoid a raucous party, one might wait until after it has settled down. Social sensors would be useful to inform people who 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 evaluate concentrations 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.

To infer the exact movement of a person using Twitter, each tweet must have accurate time and location information. Without accurate information, any inference of a person's movement would be unreliable. Regarding time information, the tweet posting time is recorded accurately in seconds. Regarding location information, few tweets can be recorded accurately from the viewpoint of confidentiality of personal information. Therefore, position estimation is necessary. A tweet includes 140 characters of textual information. Position estimation can be achieved by analyzing the contents. Thereby, one can estimate whether a tweet was posted near an event for which one wants to know the movement of people. By this simple process, one can ascertain the movement of people at the event accurately if one can accurately assess tweets posted near the target event. The position estimation here is not intended to estimate the position more accurately. It is only necessary to be able to estimate whether or not the event for which we want to know the movement of people is near this event. The ultimate goal is to determine the movement of people, especially deflation, which indicates the spreading out of people, with a social sensor. Also, it is necessary to estimate the tweet position to improve the accuracy of the deflation detection.

## 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 the 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 Zhu and Shasha [3] [4] and by Zhang and Shasha [5] examine bursting algorithms that monitor bursts efficiently over multiple window sizes. These techniques enable nearly 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.

Ohara et al. [6] propose a method for detecting the period during which burst diffusion occurs from the information diffusion series observed on social networks. They detected the burst period on Twitter, but they must build a social network to observe the spread of information related to the network.

Ebina et al. propose a method for real-time burst detection [7] [8]. The method achieves detection by inference of whether each event (each tweet posting) is a burst, or not. 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, but it remains unclear whether the burst state continues or immediately ends solely by the burst occurrence.

Endo et al. use a moving average to make full-fledged decisions [9] [10]. 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 certain posted 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: tweets with accurate location information are even fewer. Furthermore, much location information is ambiguous. Therefore, research is underway to obtain location information from tweet contents [11] [12]. 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, it is assumed that a person has location information for each city, ward, town, and village. However, such location information is insufficient for a specific event occurrence area. Therefore, we strive to identify and use more accurate position information estimation. The vocabulary included in the tweet text is analyzed. Therefore, it is not always easy to locate the event, even if it is described. 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 actually issued from the event venue. Nevertheless, removing such tweets from the overall data is not easy. One study presented a method to use Flickr data with location information as compensation for shortages of tweets with location information [13]. Some studies have proposed a method of estimating the location from a user's residential area and a method of estimating the location from the IP address of the device [14]. In the position estimation used for this study, the tweet text is analyzed in the same way as in the earlier study. However, it is characterized by estimating more accurate location information around the area where the event occurs, assuming that one already has location information for each municipality. Another feature is that the estimation accuracy is sufficient if it is useful for detecting deflation.

## 3 TARGET EVENT AND DATA

The target event for this experiment was the visit of the General Public to the Imperial Palace after Accession to the Throne on May 4, 2019. About 141,000 people visited the palace 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. Tweets with geotags analyzed for this study were sent from areas circumjacent to the Imperial Palace. Those in this range were visitors of the general public who tweeted while they waited or after they left. One 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.

We chose this event as a target because the people's movements are simple and because it is known exactly when they move. The goal is judging people's movements by analyzing tweets. Therefore, the correct answer must be found to know whether the judgment is correct. It would be interesting to know the movement of people at events such as Halloween parties and the Gion Festival. However, even if Halloween and the Gion Festival were target events of the experiment, the correct answer of people's movements could not be known. People's movements are complex: obtaining accurate information about where and when and how many people are congregating is difficult. Even if the location information were estimated, it cannot be evaluated whether the estimation is correct. By contrast, people's movements are simple during the visit of the General Public to the Imperial Palace after Accession to the Throne. Moreover the

movements are known accurately. Evaluation of the experiment results can also be done quantitatively.

The tweets to be analyzed were tweets including geotag "coordinates". The geotag "coordinate" data represent single latitude and longitude coordinates for the location at which a terminal is used for posting a tweet. Location information is obtainable from the GPS function of the terminal that posts the tweet or the access point connected to the network. It is attached to the tweet it posts. Therefore, the data are highly useful as positioning data. Tweets were extracted during 00:00:00 – 23:59:59 on May 4, 2019: the day the target event took place. The extraction range of tweets, the area, is defined by the following four latitude and longitude coordinates shown in ①–④ in Fig. 1. This is the venue for the target event and the area surrounding the front yard of the Chowa-Den where people moved.

[35.677002, 139.753658]
[35.689604, 139.753658]
[35.689604, 139.761212]
[35.677002, 139.761212]

By filtering the data using the conditions shown above, 198 tweets were extracted. Multiple tweets were posted from the same account. There were 116 unique accounts.

Their Majesties the Emperor and Empress presented their First Appearance at 10 am. Appearances were held six times until 3 pm. People move to the Chowa-Den front courtyard about 20 min before appearances, but they waited until then. Twenty minutes before the appearance, people move and waited for their appearance to hear the words of His Majesty the Emperor. When an appearance is concluded, people will disperse from the front of the Chowa-Den. People can not afford to tweet during this time. In Table 1, it is expressed as "Move". People wait about 30 min for the next appearance to start. Therefore, as Table 1 shows, we estimated how long people would have moved (or stayed).
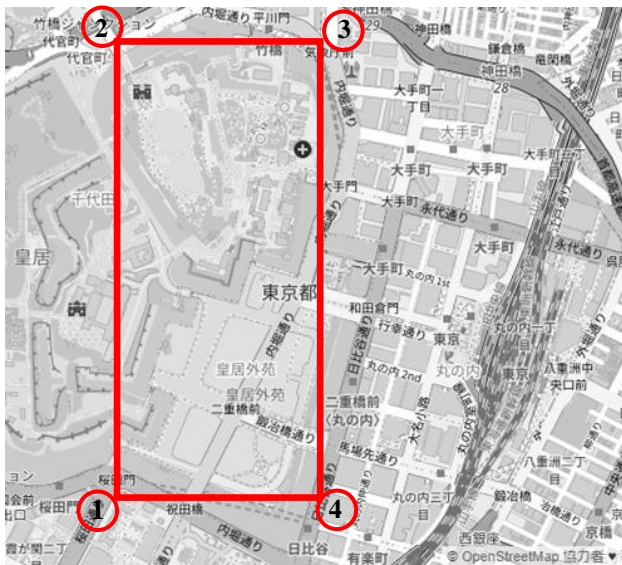


Figure 1: Target Area(in front of the Imperial Palace).

Table 1: Correct answer to estimate.

|  | time | action of people |  |
|---|---|---|---|
| before open gate | – 9:40 | Stay |  |
| 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 – | Move | Deflation |

## 4 DETERMINATION METHOD

The following two methods are used as the deflation determination method. One is 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, a person's movement is judged in units of minutes instead of days. The other method is derived from real-time burst determination as reported by Ebina et al. Our criteria are reversed to assess a deflation rather than a burst.

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

This method uses the tweet posting frequency, which is related closely to the tweet posting interval. Tweet post intervals have high real-time characteristics because they depend on each tweet. However, a certain posted interval must be found to calculate the tweet posting frequency. Usually, since the posting interval for calculating the posting frequency is set larger than the individual tweet posting interval, the real-time property of the analysis using the tweet posting frequency is lower than the analysis using the posting interval.

The method presented by Endo et al. uses a moving average of the frequency of posting tweets to estimate the full bloom of cherry blossoms or other phenomenon. 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. In this study as well, a comparative experiment was conducted using this condition. However, the tweet posting frequency is not calculated on a daily basis in this study, but on a 5-min basis. Unlike efforts to infer the best time to view cherry blossoms and so on, we wanted to assess the movement and congestion of crowds of people. Therefore, shorter intervals of 5 min were used instead of 1 day in the frequency calculation.

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

|  | 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% |

From the change in the moving average of the tweet posting frequency, one can examine how accurately the movement of the person in Table 1 can be found. Because it is a moving average, the frequency of posting tweets does not change drastically. One can use two moving averages of different lengths. For people who are actually moving, the short moving average is smaller than the long moving average. We also changed the length of the average of the moving averages to be compared. Table 2 presents results of the quantitative evaluation. Not only was judgment based on the difference between 5 moving average and 7 moving average evaluated; we also evaluated judgment results obtained when using the difference between the 3 moving average and 5 moving average and judgment results obtained when using the difference between the 3 moving average and 10 moving average. The precision is high, but the recall is low. Moving averages are used. Therefore, it is not possible to respond sensitively to changes. Moreover, there are many oversights. Both of those shortcomings engender poor recall.

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

The method of Ebina et al. 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 applied by the change of multiple tweet posting intervals. Specifically, if the moving average of the tweet interval becomes short, it is judged as a burst. Using this method, deflation is inferred by reversing the judgment conditions. In other words, the deflation occurrence condition is attained when the tweet posting interval becomes longer than before.

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% |

Similar to the discussion presented in the preceding section, we examined how accurately the deflation of the person in Table 1 can be judged under the deflation judgment condition explained above. Table 3 presents quantitative evaluation results. The change in the tweet posting interval is compared with the average of the earlier tweet posting intervals. Each row in the table is compared with the average posting interval of the 5 prior, the average posting interval up to 10 prior, and the average posting interval up to 15 prior. The recall rate of the method of Ebina et al. is high because it reacts in real time. However, its precision is not as good as that achieved when using the method reported by Endo et al. Because it has excellent real-time performance, however, an excessive reaction sensitively causes a decrease in the precision rate. It is possible to judge without overlooking the phenomenon that deflation and the recall and F-value are high.

## 5    LOCATION ESTIMATION

To improve the accuracy of deflation judgment, we will strive to increase the size and the quality of the dataset used. In the preceding chapter, we estimated people's movements using tweets with geotag "coordinates" that can provide accurate location information. However, only 116 accounts posted the tweets used in that experiment. The number of visitors involved in the general visit was 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 will perform verification by increasing the number of analysis targets using tweets with unclear positioning. We will use machine learning to extract tweets posted at the target event venue from the group of tweets including "place" that represents Chiyoda ward. We estimate the posting position. 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 visit of the General Public to the Imperial Palace from tweets that are clearly posted in Chiyoda ward.

### 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 geotag "place" has a rectangular range of information represented by four latitudes and longitudes. This rectangle delimits location information at the city level. Compared to the number of tweets that have the geotag "coordinates" as accurate location information, the number of tweets that include only the geotag "place" is larger. It depends on the location, but a difference of about four times in Japan exists overall. 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.
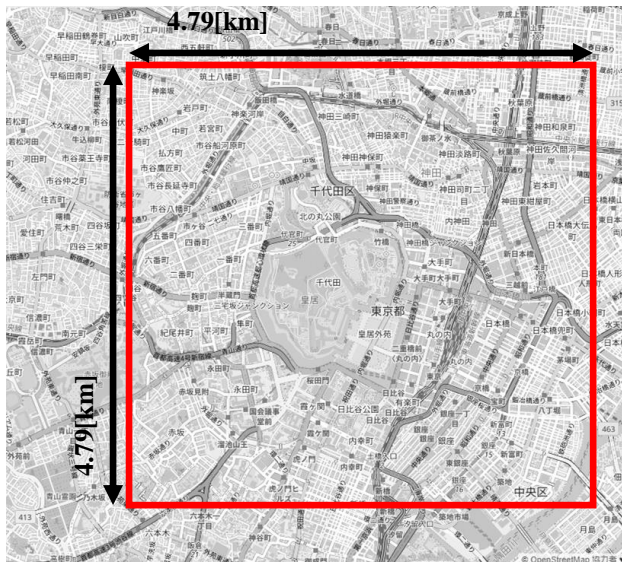
Figure 2: "place" of Chiyoda Ward.

The attached "place" data were confirmed for the 198 tweets with "coordinates" used in the preceding 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 for this place is the range entirely circumscribing Chiyoda ward.

Tweets that have only the geotag "place" and no geotag "coordinate" are shown in the range of latitudes and longitudes of the four points. They can be narrowed down to the municipality, but the exact 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 the Imperial Palace. When classifying tweets with only the geotag "place" added by machine learning, teacher data are extracted from text data of tweets with geotag "coordinates" added. For classification, the target data used for estimating the position were text data of the tweet to which only the geotag "place" representing Chiyoda ward was added. We used only tweets posted on May 4, 2019, when the target event was held.

## 5.2 Vectorization and dimension reduction

To use machine learning, we create teaching data consisting of a set of tweets as a model. These teaching data include 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 by dividing it into words. This time, we extracted only nouns. Then, assuming that tweets with the geotag "coordinates" are few, 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, after word appearances are counted for each tweet, a matrix is generated from the counted words.

The machine learning result is affected by vectorizing the extracted words using morphological analysis. A great amount of noise is included if one simply vectorizes words without consideration of their importance and meaning. Two preprocesses, TF-IDF and LSI, were applied to use the feature vector extracted from each tweet's contents 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. In addition, 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: a learning model that can execute 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 preceding chapter were used as correct answers. The tweet group in the range (including the Imperial Palace) where the event occurred is used as correct answer data.

Incorrect answer data were extracted from tweets originating from areas where the event did not occur: not including the Imperial Palace. These tweets have a "coordinates" geotag. The range of the incorrect answer data is the same area
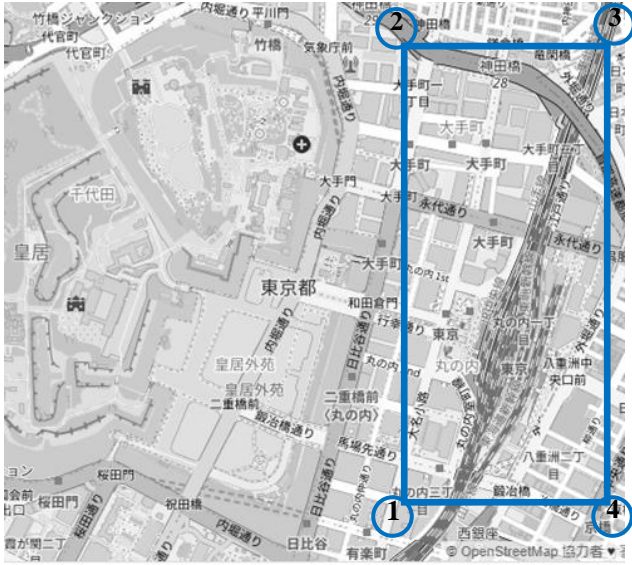
Figure 3: Area of incorrect data.

as that of the correct answer data. This is the target event location: 200 m east of the Imperial Palace. The range of latitude and longitude from the southwest is given below. Figure 3 portrays the approximate range of the extracted incorrect data (not correct data) on the map.

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

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

## 5.4 SVM

We used a classification method with Support Vector Machine (SVM), which is capable of binary classification positioning of a person who was at the scene of a specific event or incident from a message to which only the geotag "place" was added. Actually, SVM has been adopted in many earlier studies as a method to estimate user attributes and position from tweet contents. 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 data type comprises 198 tweets with geotag "coordinates" extracted in the correct answer range (Fig. 1), which is regarded as having caused people to loiter and flow because of the occurrence of events. The other data type

comprises 866 tweets with geotag "coordinates" extracted in the range (Fig. 3) where the target event (Visit of the General Public to the Imperial Palace after Accession to the Throne) has not occurred. The 1064 tweet data, which include data of these two types, 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 all. 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 a 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.

Similarly to Table 2, Table 4 presents results of quantitative evaluation using the method described by Endo et al. As in Table 2, parameters of three types are shown. In Table 2, we conducted evaluation using only 198 tweets having the geotag "coordinate". However, in Table 4, as a result of position estimation, the number of tweets used for evaluation was 1948, which is larger than those used for Table 2. The number of tweets has increased, but the precision and recall have not improved. Table 5 presents results of quantitative evaluation using the method of Ebina et al. In Table 5, the number of tweets used for evaluation is greater than in Table 3. Compared to the values presented in Table 3, the recall is not very good, but the precision is good.

We were able to increase, by nearly ten-fold, the number of tweets used to judge deflation by machine learning: from

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

|  | 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 method of Ebina et al.

|  | 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% |

Table 6: Evaluation using the method of Endo et al. for teaching data of 198 correct data and 495 incorrect data.

|  | Precision | Recall | F-value |
|---|---|---|---|
| 3(15-minute) moving average / 5(25-minute) moving average | 43.02% | 16.79% | 24.15% |
| 3(15-minute) moving average / 10(50-minute) moving average | 34.84% | 25.46% | 29.42% |
| 5(25-minute) moving average / 7(35-minute) moving average | 43.49% | 21.83% | 29.07% |

Table 7: Evaluation using the method of Ebina et al. for teaching data of 198 correct data and 495 incorrect data.

|  | Precision | Recall | F-value |
|---|---|---|---|
| 5 number analysis | 49.77% | 53.58% | 51.60% |
| 10 number analysis | 46.17% | 67.78% | 54.93% |
| 15 number analysis | 53.69% | 60.40% | 56.85% |

198 to 1948. The tweet posting position was estimated using SVM, but its classification accuracy rate in cross-validation is about 80%; it includes about 20% noise. Although the number of tweets used to judge deflation was increased by about 10 times, noise is included in that result. Therefore, the deflation judgment accuracy might or might not improve.

## 5.6 Estimated results with changed teaching data

Next, we changed the teaching data. No tweets are used as teaching data other than 198 tweets of correct answer data and 866 tweets of incorrect answer data. These are all tweets with the geotag "coordinate" which records the exact location. All of these were used as described in the preceding section, but the number of correct and incorrect data is unbalanced. Incorrect answer data are about five times as numerous as correct answer data. What we learned from correct data by machine learning is less than what we learned from incorrect data. Therefore, we randomly extracted only 495 tweets without using all incorrect answer data. This is exactly 2.5 times the number of correct data. First, a five-fold cross validation test was performed with 198 tweets as correct answer data and 495 tweets as incorrect answer data. Results show that the classification accuracy rate was 68.84%. Compared to the use of 866 tweets as incorrect answer data, the decomposition correct answer rate decreased by 10% or more.

We also investigated the case in which only 198 tweets were selected randomly from 866 tweets as incorrect answer data. The numbers of correct answer data and incorrect answer data are the same. When five-fold cross validation was applied using these 396 tweets, the decomposition correct answer rate was 90.86%, an improvement of 10% or more.

The number of teaching data and the decomposition accuracy rate are not always correlated. Moreover, the decomposition correct answer rate does not correlate with the number of incorrect answer data in the teaching data.

Five-fold cross validation was applied to a total of 693 tweets, comprising 198 tweets as correct answer data and 495 tweets as incorrect answer data. Results show that the decomposition accuracy rate was 68.84%. All of these 693 tweets were learned as teacher data. The position of 3132 tweets for which accurate position information was unknown was estimated. The geotag "coordinate" is not recorded in these 3132 tweets, but the geotag "place" is recorded. They are tweets posted in Chiyoda Ward. From position estimation by SVM, 350 tweets out of 3132 tweets were output as correct answers. It was estimated that these 350 tweets were posted within the target area of Figure 1. These 350 tweets are 548 tweets in addition to the 198 tweets for which the posting position was originally clear. With these 548 tweets, the methods of Endo et al. and of Ebina et al. were used to evaluate deflation, with results presented respectively in Tables 6 and 7.

Only 350 tweets were added by estimating the position. Compared with Table 4 and Table 5, for which 1750 tweets were added by position estimation and then deflation judgment was performed, there is little difference overall. In fact, it is slightly worse. The result is not an improvement compared to Tables 1 and 2, which show evaluation results of deflation judgment performed using only 198 tweets with the correct posting position. This can be regarded as a natural result because the decomposition correct answer rate by five-fold cross validation is also poor and data are few.

Table 8: Evaluation using the method of Endo et al. for teaching data of 198 correct data and 198 incorrect data.

|  | Precision | Recall | F-value |
|---|---|---|---|
| 3(15-minute) moving average / 5(25-minute) moving average | 49.24% | 15.75% | 23.87% |
| 3(15-minute) moving average / 10(50-minute) moving average | 42.14% | 24.13% | 30.69% |
| 5(25-minute) moving average / 7(35-minute) moving average | 44.53% | 17.61% | 25.24% |

Table 9: Evaluation using the method of Ebina et al. for a teaching data of 198 correct data and 198 incorrect data.

|  | Precision | Recall | F-value |
|---|---|---|---|
| 5 number analysis | 56.95% | 50.29% | 53.41% |
| 10 number analysis | 55.85% | 64.22% | 59.74% |
| 15 number analysis | 54.99% | 48.70% | 51.66% |

For five-fold cross validation applied with 396 tweets with the same number of correct and incorrect data, the classification correct answer rate was 90.86%. Position estimation was performed using only these 396 tweets as teacher data. By machine learning, we obtained data estimated as the correct answer for 1738 tweets from 3132 tweets that have only the geotag "place". In other words, it was estimated that 1738 out of 3132 tweets were posted in the target area presented in Fig. 1. Adding 198 tweets with known location information to these 1738 tweets yields 1936 tweets. Using these 1936 tweets, deflation judgment was performed using the methods of Endo et al. and of Ebina et al. Table 8 and Table 9 respectively present the results.

Compared with Tables 4 and 5 with deflation judgment performed by adding the 1759 tweets obtained using position estimation, it is worse overall. Furthermore, these are not good compared to Tables 6 and 7, where only 350 tweets were added by location estimation. The classification correct answer rate by five-fold cross validation yields the best result, but high accuracy of position estimation for 3132 tweets, which have only the geotag "place", is not guaranteed. The assumption that 1738 out of 3132 tweets were posted in the target area of Fig. 1 is slightly high. Tokyo Station is outside of the target area. Therefore, more tweets should be posted outside the target area.

## 6 DISCUSSION

Many data must be used to judge deflation in real time. Few tweets have geotag "coordinates" that can specify the position. Therefore, we use machine learning to estimate the locations 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 method presented by Endo et al. did not improve the judgment accuracy, but the method presented by Ebina et al. did. We have improved the accuracy rate of deflation judgment using the method of Ebina et al. by increasing the number of tweet data that can be used by estimating the location information. Results show that the recall has not improved, but the F-value has improved.

In other words, SVM location estimation increased the number of tweets posted at the target event location. The classification accuracy rate was not bad, as indicated by five-fold cross validation performed only with tweets with the geotag "coordinate". Nevertheless, because the position estimation accuracy is insufficient, when judging the movement of people using the obtained tweets, the judgment accuracy might decrease instead of increasing. Five-fold cross validation showed that tweets posted in the target area presented in Fig. 1 were used as correct answer data; tweets posted in the area depicted in Fig. 3 were used as incorrect answer data. These areas are not adjacent and might have contributed to a better classification accuracy rate. It might be true that tweets posted in the area presented in Fig. 2 excluding the tweets posted in the area presented in Fig. 1 should have been regarded as incorrect data. In that case, if those tweets were used as teaching data for machine learning, the tweets posted near the target area in Fig. 1 would become noise and therefore reduce the classification accuracy rate. This point must also be investigated carefully.

Although SVM was adopted for machine learning, various other methods are available. Future studies must compare results obtained from their use. For this study, the data of the tweet contents learned by machine learning were only nouns. Verbs, because of conjugation variants, were avoided. As one might expect, they are difficult to handle. It is also worth considering how verbs and adjectives should be used, in addition to word-dependence. The use of information such as accounts, rather than the tweet body text, should also be considered. Additionally, it is necessary to consider methods other than judging the movement of people as a method of using tweets for which the position is estimated. Various target events can be considered when judging the movements of people according to similar stimuli. These points will be addressed in future research.

## 7 CONCLUSION

Numerous data must be accumulated to support real-time judgment of the movement of people, especially for deflation, whereby people disappear and spread out. Few tweets include accurate location information. Therefore, for a given area, we used machine learning to estimate the locations of tweets that have a "place" geotag, which is ambiguous location information. Deflation was inferred from tweets that were presumably posted at a certain place. Because of imperfect estimation of position information, the deflation judgment accuracy did not necessarily improve. However, the deflation judgment accuracy improved in some cases.

Some room exists for improvement of position estimation. Future tasks are consistent and reliable improvement of deflation judgment accuracy.

# 8 ACKNOWLEDGMENTS

# 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, Vol. 27, No. 1, pp.67-74 (2012).

[2] J. Kleinberg, Bursty and hierarchical structure in streams, In Proc. Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp.91-101(2002).

[3] Y. Zhu, and D.Shasha, Efficient Elastic Burst Detection in Data Streams, Proc. Ninth ACM SIGMOD International Conference on Knowledge Discovery and Data Mining, ACM, pp.336-345(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, IEEE Computer Society, pp.146-149 (2006).

[6] K. Ohara, K. Saito, M. Kimura, and H. Motoda: "Burst Detection in a Sequence of Tweets based on Information Diffusion Model", DBSJ journal, Vol.11, No. 2, pp.25-30(2012).

[7] 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).

[8] R. Ebina, K. Nakamura, and S. Oyanagi, A Proposal for a Real-time Burst Analysis Method, IPSJ TOD, Vol.5, No.3, pp.86-96(2012).

[9] 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).

[10] 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).

[11] 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).

[12] 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).

[13] T. Mansouri, AZ. Ravasan, and MR Gholamian, A novel hybrid algorithm based on k-means and evolutionary computations for real time clustering, Int J Data Warehous Min (IJDWM) 10(3):1–14 (2014).

[14] J. Eisenstein, B. O'Connor, NA. Smith, and EP. Xing, A latent variable model for geographic lexical variation, In Proceedings of the 2010 conference on empirical methods in natural language processing, Association for Computational Linguistics, pp 1277–1287(2010)

**Takuma Toyoshima** is currently working as vocational training instructor in a course of training that production support with ICT at the Polytechnic center Shiga. He received a Master of degree in Industrial Engineering from the NIAD-QE Tokyo in 2020.

**Takuo Kikuchi** received the Ph.D degree in engineering from Kyushu Institute of Technology. He is currently a professor, dept. of electrical and computer engineering at Polytechnic University. His research interests include "Information network cabling" and "Skill analysis and evaluations". He is a member of the JSEE, IEICE and JSET Japan.

**Masaki Endo** earned a B.E. degree from Polytechnic University, Tokyo and graduated from the course of Electrical Engineering and Computer Science, Graduate School of Engineering Polytechnic University. He received an M.E. degree from NIAD-UE, Tokyo. He earned a Ph.D. Degree in Engineering from Tokyo Metropolitan University in 2016. He is currently an Associate Professor of Polytechnic University, Tokyo. His research interests include web services and web mining. He is also a member of DBSJ, NPO STI, IPSJ, and IEICE.

**Shigeyoshi Ohno** earned a M.Sci. and a Dr. Sci. degrees from Kanazawa University and a Dr. Eng. degree from Tokyo Metropolitan University. He is currently a full Professor of Polytechnic University, Tokyo. His research interests include big data and web mining. He is a member of DBSJ, IPSJ, IEICE and JPS.

**Hiroshi Ishikawa** earned B.S. and Ph.D. degrees in Information Science from The University of Tokyo. After working for Fujitsu Laboratories and becoming a full Professor at Shizuoka University, he became a full Professor at Tokyo Metropolitan University in April, 2013. His research interests include databases, data mining, and social big data. He has published actively in international refereed journals and conferences such as ACM TODS, IEEE TKDE, VLDB, IEEE ICDE, and ACM SIGSPATIAL. He has authored several books: Social Big Data Mining (CRC Press). He is a fellow of IPSJ and IEICE and is a member of ACM and IEEE.

# Development of Yield Prediction Model Generation Process for Fruit Vegetables in Plant Factories

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, takamori@agricc.biz

*Abstract* – Plant factories in Japan have become increasingly popular in recent years. In these factories, data are collected using production management systems and both external and internal environmental sensors. However, predicting yields for fruit and vegetable crops is difficult because these crops have unique biological characteristics, and their growth depends on weather conditions. Thus, the purpose of this study is to develop a yield model generation process for fruits and vegetables in plant factories. By defining a process for generating predictive models, we aim to improve the efficiency and accuracy of their development, as well as to make them applicable to various fruit and vegetable crops and to various facilities. This paper reports on the results of our application of the proposed predictive model generation process to develop a model for mini cucumbers and mini tomatoes, which we interpreted as being representative of other fruit and vegetable crops. Experimental results show that the proposed model generation process can be applied to various crops. In addition, it was confirmed that the tsfresh Python package, which we used to automatically extract features from time-series data, improved the prediction accuracy.

*Keywords*: Yield Prediction, Plant Factory, Fruit Type Vegetables, Statistical Modeling, tsfresh

## 1 INTRODUCTION

In recent years, there has been great interest in the promising field of protected horticulture, including next-generation horticulture. This new approach involves equipping sites with advanced environmental control devices utilizing information and communications technology (ICT) and other technologies, aiming to integrate facilities and efficiently use local resources and energy [1]. To extend the shipping period for vegetables, horticulture facilities in Japan have been upgraded from plastic tunnels and rain shelters to greenhouses and then to plant factories with highly controlled greenhouse environments.

Plant factories are a form of horticultural agriculture that enables year-round, planned production of vegetables and other plants through advanced environmental control and growth forecasting [2]. They collect and accumulate large quantities of data using production management systems and environmental sensor data inside and outside the factory. Plant factories can be broadly classified into two types: (1) the "sunlight-based" type, in which plants are cultivated in

greenhouses, or similar structures using sunlight, supplemented by artificial light and technology to prevent high temperatures in summer and (2) the "fully artificial light-based" type, in which plants are cultivated in a closed environment without sunlight. The main crops grown in the artificial light type are leafy greens, such as spinach and lettuce, and fruit crops such as tomatoes and cucumbers are commonly grown in facilities of the sunlight type. Plant factories in Japan are still developing, and there are still major problems such as low yield per farm area [3][4] and low labor productivity [5][6]. Therefore, it is necessary to establish a new Japanese-style cultivation platform, standardize facilities, and promote research benefiting the average of cultivation expertise.

One of the challenges faced by plant factories is the difficulty of predicting crop yield [7][8]. Because fruit vegetables are mainly grown in sunlight-type plant factories, they are subject to variations in sunlight, temperature, and other seasonal weather conditions, significantly affecting their growth. Given these factors and because of their biological characteristics, yield prediction for fruit vegetables is difficult. Compared with leafy vegetables, fruit vegetables are more affected by environmental conditions for a longer duration as they have both flowering and fruiting periods. Moreover, as fruit vegetables can be harvested from a single seedling several times, predicting their yield is a challenge. Yield prediction is important to match market demand and prevent overproduction.

Yield prediction and the relationship between yield and environmental factors have been extensively studied for crops grown in open fields or greenhouses with simple environmental controls and major crops such as paddy and wheat. However, we have not found any studies on fruit and vegetable crops grown in plant factories.

Therefore, the purpose of this study is to propose a model generation process for fruit vegetables grown in plant factories. The model generation process comprised six sequential steps: (1) data selection for model generation, (2) data preprocessing, (3) data visualization, (4) feature design, (5) selection of prediction methods, and (6) model optimization. By defining the process for generating predictive models, we aim to improve the efficiency and accuracy of their development as well as to make them applicable to various fruit vegetable crops and other facilities.

## 2   RELATED RESEARCH

### 2.1     Prediction Method

Many studies have investigated yield prediction using various machine learning techniques, such as artificial neural networks and boosted regression trees [9]-[12]. Such techniques have achieved high prediction accuracy; however, machine learning requires a large amount of training data, which limits its application.

Apart from machine learning, statistical modeling has also been used as a prediction method. Related studies include those by Hoshi et al. [7] and Okuno et al. [13]. In 2000, Hoshi et al. predicted the daily yields of tomatoes grown in a greenhouse using topological case-based modeling (TCBM) [14] and multiple regression analysis. TCBM 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 generated a regression model that can be effectively used as a sound basis for prediction results and estimated yields by capturing increasing and decreasing trends. However, there is insufficient information on the actual operation due to the limited number of prediction methods and crops evaluated.

### 2.2     Feature Generation

In model development, it is important to consider the features to be incorporated. Many of the abovementioned related studies incorporated features related to environmental data, such as temperature, humidity, light, and precipitation. In general, only the most basic statistics, such as average, maximum, and minimum values over a period of time, were used to generate features in the relevant literature. However, this method has the drawback that finding factors related to yield is difficult because of the small variation in features, and the accuracy of the generated models is limited.

### 2.3     Research Tasks

As mentioned previously, the model generation process formulated in this study consisted of six subprocesses, performed in the following order: (1) data selection for model generation, (2) data preprocessing, (3) data visualization, (4) feature design, (5) selection of prediction methods, and (6) model optimization. Based on the issues discussed in related studies, designing the features and selecting a prediction method are the major challenges. Therefore, the two research tasks in this study were:

(1)  Selection of a method for building a predictive model

Preparing a large dataset is difficult for plant factories due to significant environmental and cultivar variations caused by the equipment. Therefore, it is necessary to build a predictive model that can be used even with a relatively small amount 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 plant factory employees are able to clearly understand in practical use because unconvincing prediction models are generally unacceptable to farmers.

(2) Selection of feature extraction method

Actual yield data are 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 take this into account. Thus, it is necessary to calculate various features, such as median, variance, standard deviation, Fourier transform, and autocorrelation coefficients.

## 3   PROPOSED METHOD

### 3.1     Approaches

The approaches employed to carry out the research tasks described in Section 2.3 are as follows.
(1) We selected a prediction model that works even with a small amount of data and has a comprehensible model structure.
(2) We also selected methods for extracting various features from time-series data.
A more detailed description of these approaches is given in Section 3.2. The model generation process that incorporates all these approaches is also described.

### 3.2     Model Generation Process

As mentioned, the model generation process involves six subprocesses. In this research, these processes were formulated as simple arithmetic procedures to identify factors related to yield and generate a model that can accurately predict yield according to various facility-specific conditions. In each subprocess, a representative fruit was used to develop the model. For subprocess (1) to (6), we show the process of searching for an optimal model generation process for the clauses corresponding to the numbers. Figure 1 shows the algorithm of the yield prediction model generation process, which is the proposed method.

#### 3.2.1   Data Selection

In this study, we decided to use three major types of data as candidate features of two crops: actual data of past yields, environmental data in the facility, as well as outdoor weather data. There are two main reasons for this.

The first reason is that the data are continuously collected and recorded in plant factories. Some related studies have used data on crop appearances, such as the normalized difference vegetation index (NDVI) and crop stem diameter measurement data [9]-[12]. However, this method requires the installation of new equipment for sensing the appearance data, which results in high data collection costs. On the other hand, environmental data inside and outside structures and actual yield data are generally obtained and recorded in plant factories, thus collecting this data is easy.

```
1  // The proposed overall process and algorism(Except "selection
       for model generation" of sub process (1) )
2  // df: Data set of both target and potential predictor attributes
3  // envDf: Environmental data of data set
4  // algorithms: A list of algorithms (MLR,GAM,MARS)
5
6  procedure proposedProcess(df,algorithms){
7      // Pre-processing of environmental data
8      accumulationPeriodto7days(envDf)
9      // Completing missing values
10     completionMissingValue(envDf)
11     // Replace with preprocessed attributes
12     df = replace(envDf)
13
14     // Data visualization
15     pairplot(df)
16     heatmapOfCorrelation(df)
17
18     // Standardization
19     df = standardZScore(df)
20     // Split data frame
21     target = dfOfTarget
22     feat = dfOfFeat
23
24     // Extract features by tsfresh
25     featTs = extractFeatures(feat)
26     // Replace NaN and infinity values(tsfresh's function)
27     featTs = replaceNanInf(featTs)
28     // Statistical hypothesis testing(tsfresh's function)
29     featTs = selectByHypothesisTesting(featTs, target)
30     // Select features by mutual features
31     featTs = selectByMutualInfo(featTs, target)
32
33     // Integrate feature and objective variables
34     df = Integration(featureTs,target)
35     // Split 75% into training data and 25% into test data
36     train = dfOfFourThreeQuarters
37     test = dfOfFourOneQuarters
38
39     // Makes a model with the algorithms
40     if(MLRandGAMAlgorithm)
41         model = calculateBestAIC(algorithms,train)
42     else
43         model = makeModel(algorithms,train)
44     end if
45
46     // Detecting Multicollinearity with VIF
47     vif = VarianceInflationFactor(featTs)
48
49     // Evaluate the predictive model using a test sample
50     if(vif<5)
51        pred = evalModel(model,test)
52     end if
53  }
```

Figure 1: Algorithm of the proposed method

The second reason is that these three types of data were reported to be effective as features in a related study that predicted the yield of fruit and vegetable crops in greenhouses [7][9][14]. Greenhouses and plant factories differ in terms of environmental control methods, but they share the primary characteristic of growing crops "in the facility." Therefore, we use these attributes because we believe they are likely to be effective for fruit and vegetable crops in plant factories. We also use the weekly yield  as the objective variable and the "yield one week ago" relative to this objective variable as an attribute related to the data of past actual yields. We adopted this approach based on the results of a study previously conducted by the author [15], in which a single correlation analysis of weekly yield and past actual yield data showed that the yield one week previous had the highest correlation with the objective variable. Because multicollinearity may occur if multiple similar attributes of past yields are included and it is necessary to reduce the dimensionality of the features, we decided to use only the yield of one week before based on the results of the previous correlation analysis.

For these two reasons, environmental data inside and outside the greenhouse and past actual yield data were used.

### 3.2.2　Data Preprocessing

Data preprocessing consisted of three major components.

First, we process the integration period of environmental data for the attributes, created by using the accumulated value from 7 day prior to the integration up to the morning of the reference date (the day before the forecast date). This period corresponded to the fruiting period, which is the period from flowering to harvest for both mini cucumbers and mini tomatoes. As environmental factors during this period have been shown to have a significant impact on yields [16], the data during this period were used. In addition, we create attributes for daytime hours only as plants grow mainly when there is sunlight for photosynthesis. Because the times of sunrise and sunset change throughout the year, we use the annual average of the time period of daylight in the region where the facility is located.

The second component is the replacement of the missing values. Missing values due to malfunctioning sensor devices were supplemented by the average value for the three days before and after the missing value occurred.

The third component is the integration of data scales. To do this, we standardize the process using a robust z-score. Robust z-scores were chosen as their median is not affected by the shape of the distribution, and the quartiles are not affected by the outliers at both ends of the distribution (statistic of variability).

### 3.2.3　Data Visualization

Data visualization facilitates the analysis of relationships between objective variables and features and identifies outliers. Scatter plots can be used to visualize the data. A scatter plot takes one feature on the x-axis and another on the y-axis and plots dots at each data point. In this study, as there were more than three features, a scatter plot matrix (paired plot) was used to plot all possible feature combinations. In addition, to obtain a clearer picture of the correlations between the attributes, a heat map of the correlation matrix, as well as a scatter plot diagram, was used.

### 3.2.4　Feature Design

The design of features involves two steps: extraction and selection.

First, we discuss the feature extraction method. Various features are generated by extracting the features of time series data, as described in Section 3.1. In time series data, the observed values and the times of observation are recorded, and it is necessary to generate features that capture the characteristics of the order of the data and the backward/forward relationship. To extract various features specific to time series data, we used a suitable Python package called tsfresh (Time Series FeatuRe Extraction on the basis of Scalable Hypothesis tests) [17],  which includes feature extraction and feature selection algorithms for time series analysis. Figure 2 shows a schematic of our process flow using this package.  tsfresh provides 63 characterization methods for the extraction of 794 distinct features. For example, there are mean, maximum, minimum, number of peaks, median, standard deviation, and

Fourier transform features, as well as features using autocorrelation coefficients and time-reversal symmetry features. We used tsfresh because it can generate a comprehensive set of features specific to time series data, and the extraction and selection processes can be parallelized to significantly reduce execution time. In addition, tsfresh has been used in many research papers and in various applications such as disease prediction, machine fault detection, and traffic volume prediction.

Next, we describe four methods of feature selection.

The first method uses the select_features module provided by tsfresh, which selects features using statistical hypothesis testing so that only features that are likely to have statistically significant differences are selected. The significance testing methods for features included Fisher's exact test of independence, the Kolmogorov–Smirnov (KS) test for binary features, the KS test for continuous features, and a Kendall rank correlation coefficient test, depending on the type of supervised machine learning problem (classification or regression) and the type of feature (categorical or continuous) [18]. The result of validation by these methods is a vector of p-values, which quantifies the importance of each feature for predicting labels and targets. In Fig. 2, this corresponds to the third stage, "p-values." These p-values were evaluated based on the Benjamini–Yekutieli procedure to determine which features to retain. This stage corresponds to the last stage, "Selected features," in Fig. 2.

The second method is based on mutual information (MI). MI calculates the dependence of the product of the simultaneous distribution P(X,Y) and the individual distributions P(X)P(Y) between one feature X and another feature Y. If they are independent of each other, MI is zero. According to the MI, the number of features should be chosen to be "the number of samples in the training data/10." This is d to the amount of training data is not appropriate in terms of noise pattern learning and learning speed.

The third method is the stepwise method based on the Akaike Information Criterion (AIC), which is a common statistical variable selection method. The AIC is used to find the best combination of features in the prediction models of multiple regression analysis and generalized additive models, which will be explained in the next section. The multivariate adaptive regression spline, which is another prediction because having too many features compare method used, automatically performs feature selection in the algorithm, so this feature selection method is applied to both multiple regression analysis and generalized additive models

In the fourth method, we use the variance inflation factor (VIF) to check whether multicollinearity occurs. If multicollinearity is present, the corresponding variable is deleted and the feature is selected.

The mentioned above, we use four methods of feature selection.

### 3.2.5    Prediction Method Selection

Our methods, requiring only a small amount of data and involving a comprehensible model structure, are described in this section. The selection criteria for the prediction methods
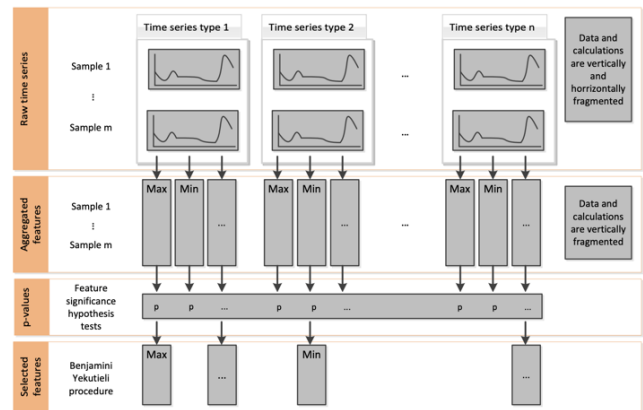


Figure 2: Data processing flow from feature extraction to selection (adapted from [18])

are as follows: (i) the regression structure of the features can be searched and evaluated, and (ii) a model with high prediction accuracy and low data requirement can be constructed. Three modeling methods that satisfied both criteria were selected for the study.

The first method was 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, including fruits and vegetables [7][13][16]. MLR was selected because it has some degree of prediction accuracy. To determine the best feature combination, a stepwise method based on the Akaike information criterion (AIC), a common statistical variable selection method, was implemented in our MLR approach. The second method used was the generalized additive model (GAM), selected because it can predict with an accuracy similar to that of a machine learning model. It also retains the advantage of a linear model, where the relationship between objective variables and features can be easily determined. Similar to MLR, AIC was also implemented in GAM. The third method used was multivariate adaptive regression splines (MARS). Compared to GLM, MARS can explicitly represent the interaction between features, including tipping points in tree structures [14].

### 3.2.6    Model Optimization

The target data were divided into training data (75 %) and test data (25 %). The datasets were evaluated using the leave-one-out method. The correlation coefficient (R) and the mean absolute error (MAE) were 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. As percent yield varies among crops, MAE was expressed as a percentage relative to the average yield.

We evaluated the prediction accuracy of the developed model using R and MAE, and the method with the highest prediction accuracy was selected.

# 4 EXPERIMENTS

## 4.1 Target Facility

The experimental facility used in this study was 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 sells hydroponically grown fruits and vegetables (7 fruits and 17 leafy vegetables). Apart from collecting environmental data, such as temperature, humidity, $CO_2$ concentration, and nutrient concentration in the hydroponic solution, Plant Factory A collects external weather data, such as temperature, humidity, precipitation, and light intensity. Monitoring both external and internal environmental parameters is useful for controlling the environment of 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 Crops

We used mini cucumber (Larino White) as the base case, and mini tomato (Aiko) as the target crop to verify the versatility of the proposed forecasting model generation process. These two crops are among the top three major crops cultivated in horticulture facilities in Japan. Because they are also the main crops grown in Plant Factory A, their cultivation area is large, and they are shipped almost every day. Therefore, there is a high demand for the development of a yield prediction model for these crops.

## 4.3 Target Data and Problem Setting

In this subsection, we describe the details of the data used to construct the model and the problem setting based on interviews with plant factory employees.

### 4.3.1 Target Data and Preprocessing

The data used for model building and evaluation were all 66 weeks of the period from February 2018 to June 2019. The selection of attributes was based on the approach described in Section 3.2.1. Table 1 lists the candidate attributes for the feature values. As explained in Section 3.2.2, preprocessing such as changing the integration period of the environmental data was performed for the attributes in Table 1.

### 4.3.2 Problem Setting

From the interview with the management and production employees, we set the yield for a single week from the day following the day on which the yield was predicted (hereafter, the prediction date) as the objective variable, hereafter referred to as the weekly yield and predicted in this study using environmental and production data up to the forecast date. Plant Factory A had not been able to predict the yield of fruit and vegetable crops at all, so it is necessary to bring them to the stage of utilization as soon as possible. Based on these

Table 1: Potential predictor attributes

| Attribute code name | Attribute name and description |
|---|---|
| YW | Yield weight (kg) |
| PYW | Past yield weight a week ago (kg) |
| AT | Air temperature (°C) |
| AH | Air humidity (%) |
| SR | Solar radiation (kWh m$^{-2}$) |
| FT | Temperature in the facility (°C) |
| FCD | Carbon dioxide concentration in the facility (ppm) |
| FS | Saturation humidity in the facility, in free water vapor capacity per 1 m$^3$ of air (g/m$^3$) |

circumstances and the general accuracy standards for the practical use of prediction models, we aim to achieve a prediction accuracy of R = 0.8 or better and MAE = 20.0 % or less.

## 4.4 Data Visualization

### 4.4.1 Visualization of Weekly Yield Trends

Figure 3 shows a graph of the weekly yields of the predicted targets. The average weekly yield was 135 kg for mini cucumbers and 24.5 kg for mini tomatoes. Compared to mini tomatoes, mini cucumbers showed a larger variation in yield. Trend pattern analysis of the time series data showed that the yield of mini cucumbers fluctuated seasonally. In Fig. 3, the period of large increase in the yield of mini cucumbers corresponds to the summer season, indicating that the yield is at its maximum during the summer when the light intensity increases.

On the other hand, the yield of mini tomatoes showed cyclic variation and only a slight trend of seasonal variation. Fig. 3 shows that there were several cycles of large yield increase and decrease, which may be due to the biological characteristics of fruit crops: new fruits are produced after a period of time. To investigate the reason for the lack of seasonal variation in the tomatoes, we interviewed the employees of the plant factory and found that they had been neglecting the cultivation management of the tomatoes because they had been focusing on the cultivation of other crops during the period. Specifically, they left some plants long after they should have been discarded, and were late in responding to pest damage.
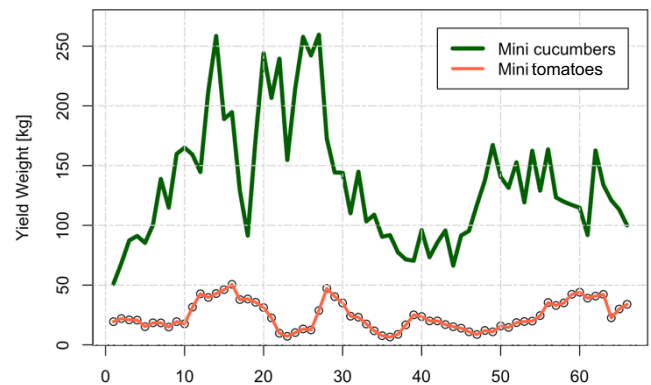


Figure 3: Weekly yield of the two crops

These results suggest that when cultivation management is properly carried out, the tendency of seasonal variation is such that the yield is highest in the summer season, as in the case of mini cucumbers.

### 4.4.2 Visualization of Attribute Relationships

The relationships between the objective variables and features and are described in this section. Figures 4 and 5 show a heat map table of the correlation coefficients. Spearman's rank correlation coefficient was used to calculate the single correlation coefficient.

Eight attributes, including the objective variables, were used in the analysis. Preprocessing was carried out, such as changing the integration period of environmental data, processing of missing values, and standardization. Table 2 lists the attributes created for both crops.

The strongest correlation for mini cucumbers was found for the weekly yield of the previous week with a single correlation coefficient of 0.79, followed by carbon dioxide concentration ($-0.71$) and air temperature (0.69). The strongest correlation for mini tomatoes was found for the weekly yield of the previous week, with a single correlation coefficient of 0.83, followed by carbon dioxide concentration ($-0.34$) and air temperature (0.33).

The same attributes were found to be strongly correlated with yield in both mini cucumbers and mini tomatoes. By adding these highly correlated attributes into the predictive model, the accuracy of the model would most likely be improved. Based on the values of the correlation coefficients for each attribute, there were significant differences between the two crops. Therefore, we inferred that the degree of influence of environmental factors varies with the crop.



Figure 5: Heat map table
of correlation coefficients for mini tomatoes
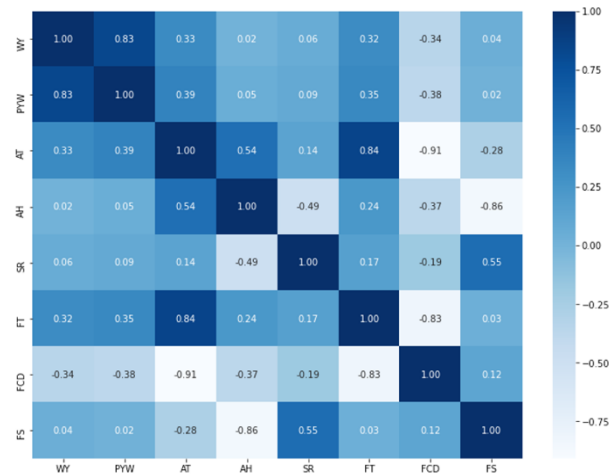


Figure 4: Heat map table
of correlation coefficients for mini cucumbers

## 4.5 Feature Design

First, we used the extract_features module of tsfresh to extract features reflecting the characteristics of the time series data from the attributes in Table 1. As a result, a combined total of 5278 features were extracted for mini cucumbers and mini tomatoes. Next, by statistical hypothesis testing using the tsfresh select_features module, we narrowed down the number of features to 82 for mini cucumbers and 69 for mini tomatoes. Next, using mutual features, we selected five features each for mini cucumber and mini tomato. Finally, the VIF values were calculated, and because all the features were less than 5, the suspicion of multicollinearity was low, and we decided to use all five features.

The features thus obtained are shown in Tables 2 and 3. The feature names are composed of the following three elements: (1) the time series attribute from which the feature is extracted, (2) the name of the feature calculator used to extract the feature, and (3) the key-value pairs of the parameters that make up the feature calculator:

[kind] _ [calculator] _ [parameterA] _ [valueA] _ [parameterB] _ [valueB]

For example, the feature name carbon-DioxideConcentration _cwt_coefficients_widths_(2, 5, 10, 20)_coeff_0_w_10 for mini cucumbers in Table 2 has kind = carbon-DioxideConcentration, calculator = cwt_coefficient, parameter width = (2, 5, 10, 20), parameter coeff = 0, and parameter w = 10, which represents the continuous wavelet transform of the Ricker wavelet of the yield of one week ago (PYW).

Next, we describe the selected features. First, we detail the features that reflect the characteristics unique to time series data. These are C_CDC_TS, C_SR_T, T_PYW_TS, T_HT_TS, and T-CDC_TS, which use the one-dimensional discrete Fourier transform and the Ricker wavelet continuous wavelet transform, shown in Tables 2 and 3. These features represent periodic changes. This periodicity was observed in both crops as seasonal and cyclic variation during the analysis

Table 2: Mini cucumber features extracted by tsfresh

| Notation | Feature name | Feature description |
|---|---|---|
| C_PYW_TS | pastYieldWeight_average_period_7 | Cumulative value of past yields from 7 days before the forecast date |
| C_AT_TS | airTemp_average_period_7 | Cumulative value of air temperature from 7 days before the forecast date |
| C_CDC_TS | carbonDioxideConcentration _cwt_coefficients_widths_(2, 5, 10, 20)_coeff_0_w_10 | Previous week's yield of the objective variable, Calculates a continuous wavelet transform for the Ricker wavelet, also known as the Mexican hat wavelet |
| C_SR_TS | solarRadiation_fft_coefficient_coeff_0_attr_angle | Fourier coefficients of the one-dimensional discrete fast Fourier transform about solar radiation |
| C_HT_TS | houseTempAve_average_period_7 | Cumulative value of temperature in the facility from 7 days before the forecast date |

Table 3: Mini tomato features extracted by tsfresh

| Notation | Feature name | Feature description |
|---|---|---|
| T_PYW_TS | pastYieldWeight_cwt_coefficients_widths_(2, 5, 10, 20)_coeff_0_w_2 | Previous week's yield of the objective variable; calculates a continuous wavelet transform for the Ricker wavelet |
| T_AT_TS | airTemp_average_period_7 | Cumulative value of air temperature from 7 days before the forecast date |
| T_SR_TS | solarRadiation_abs_energy | Sum over the squared values of solar radiation |
| T_HT_TS | houseTemp_fft_coefficient_coeff_0_attr_angle | Fourier coefficients of the one-dimensional discrete fast Fourier transform about temperature in the house |
| T_CDC_TS | carbonDioxideConcentration_fft_coefficient_coeff_0_attr_angle | Fourier coefficients of the one-dimensional discrete fast Fourier transform about carbon dioxide concentration |

of weekly yield trends, which explains the selection of these features. Other features selected were C_PYW_TS, C_AT_TS, C_HT_TS, and T_AT_TS, which use cumulative values from 7 day before the forecast date, and T_SR_TS, which represents the addition of squared values.

Subsequent model tuning and selection of the best prediction method were performed using these features.

## 4.6 Model Optimization

Of the 66 available data, 50 (about 75 %) were used as training data and 16 (about 25 %) were used as test data (Fig. 6).

First, MLR, GAM, and MARS methods were trained on the data. We searched for the best combination of features and hyperparameters in the data. Tables 4 and 5 show the combinations of features and hyperparameters with the highest accuracy. Tables 4 and 5 summarize the results of the most accurate feature combinations and optimal hyperparameters. The feature values selected differed depending on the target crop and prediction method. For each feature, the attributes that were found to have high correlation with yield in the correlation analysis tended to be selected more frequently.



Figure 6: Data partitioning method

Table 4: Model tuning results (feature selection)

| Crop | Feature | | |
|---|---|---|---|
| | MLR | GAM | MARS |
| Cucumber | C_PYW_TS | C_PYW_TS | C_PYW_TS |
| | C_CDC_TS | C_CDC_TS | |
| Tomato | T_PYW_TS | T_PYW_TS | T_PYW_TS |
| | T_SR_TS | | T_SR_TS |
| | T_CDC_TS | | |

Table 6 shows the prediction accuracy results for the test data using the tuned features and hyperparameters. For mini cucumbers, the highest prediction accuracy was achieved by MARS, and for mini tomatoes, it was achieved by GAM.

To evaluate the effectiveness of tsfresh, we compared the results of feature extraction using tsfresh and the results of feature extraction using the features in Table 2 without tsfresh. The prediction method that showed the highest accuracy for each crop was used for comparison. The results are shown in Tables 7 and 8. As a result, we confirmed that using tsfresh improved the prediction accuracy, except for the MAE in the test data for mini cucumbers and mini tomatoes

## 4.7 Discussion

For both mini cucumbers and mini tomatoes, the highest prediction accuracy was achieved when only the past yields (C_PYW_TS, T_PYW_TS) were considered. Conversely, it was found that incorporating environmental data into the features significantly reduced the accuracy. The results of this

Table 6: Evaluation results for each accuracy index

| Crop | R | | | MAE(%) | | |
|---|---|---|---|---|---|---|
| | MLR | GAM | MARS | MLR | GAM | MARS |
| Cucumber | -0.267 | 0.067 | **0.279** | 43.1 | 46.9 | **23.8** |
| Tomato | 0.755 | **0.790** | 0.720 | 42.0 | **19.8** | 42.6 |

Table 5: Model tuning results (hyperparameter)

| Crop | Hyperparameter | | |
|---|---|---|---|
| | MLR | GAM | MARS |
| Cucumber | none | Select = TRUE Method = GCV.Cp | degree=1 nprune=2 |
| Tomato | none | Select = TRUE Method = REML | degree=1 nprune=12 |

Table 7: Accuracies with and without using TSFRESH in feature extraction （mini cucumber）

| | R | | MAE (%) | |
|---|---|---|---|---|
| | Train | Test | Train | Test |
| Using tsfresh | **0.796** | **0.279** | **20.7** | 23.8 |
| Without tsfresh | 0.762 | 0.156 | 22.3 | **17.3** |

Table 8: Accuracies with and without using TSFRESH in feature extraction (mini tomato)

| | R | | MAE (%) | |
|---|---|---|---|---|
| | Train | Test | Train | Test |
| Using tsfresh | **0.851** | **0.790** | 25.7 | **19.8** |
| Without tsfresh | 0.838 | 0.770 | **19.7** | 19.9 |

experiment showed that past yields had the highest contribution to the prediction of crop yields in plant factories and that environmental data were probably noise. Environmental data made a significant contribution to yield predictions in open-field crops [8]-[13]. On the other hand, this tendency was not observed in fruit and vegetable crops in plant factories, which may be because plant factories are subject to various influences other than the environment. As shown in the visualization of the yield trend of mini tomatoes in Section 4.4.1, the influence of human factors such as production management methods and shipping adjustment methods is particularly high. In addition, we received similar comments from the employees of the plant factory. In the future, we will try to improve the accuracy of the prediction model by adding features related to human factors.

In terms of accuracy, the prediction accuracy of mini tomatoes was R = 0.790, MAE = 19.8, which was higher than that of mini cucumbers. This accuracy generally met the accuracy targets of R = 0.8 or higher and MAE = 20.0% as described in 4.3.2. This result indicates that the proposed model generation process is applicable to other crops. On the other hand, the accuracy for mini cucumbers was R = 0.279 and MAE = 17.3. MAE reached the target value, but R did not reach the target value and was lower than 0.300, which is generally

considered to be correlated. This difference in accuracy is largely due to the variation in average yield. The average yield of mini cucumbers was 135 kg, whereas that of mini tomatoes was 24.5 kg, a difference of more than five times. Therefore, the error value is more likely to be larger for mini cucumbers. In the future, it will be necessary to improve the prediction model generation process so that a certain level of accuracy can be achieved even in the case of crops with large fluctuations in average yield and weekly actual yield, such as mini cucumbers.

We asked the employees at the plant factory to confirm the accuracy of the test results, and they said that the yield prediction accuracy for the mini tomatoes was practically viable, but using the approach for mini cucumbers was not yet practicable with the current accuracy. They pointed out that to achieve a practical level, it is more important to know the increasing or decreasing trend of yield compared with the prediction error. In the future, we will improve the accuracy with the goal of identifying yield trends to meet the needs of employees in the plant factory.

# 5   CONCLUSION

## 5.1 Summary

The purpose of this study is to develop a model generation process for fruit vegetables in plant factories. The aim is to improve the efficiency and accuracy of plant factory crop yield prediction models and to make them applicable to various fruit and vegetable crops as well as facilities by having a defined process for model generation. The model generation process comprises (1) data selection, (2) data preprocessing, (3) data visualization, (4) feature design, (5) selection of prediction methods, and (6) model optimization. This paper reports on the results of applying the proposed predictive model generation process to mini cucumbers and mini tomatoes, which we interpreted as being representative of other fruit and vegetable crops. Application of the proposed model generation process to mini tomatoes resulted in R = 0.790 and MAE = 19.8, indicating that the proposed method is applicable to other crops. In addition, it was found that the tsfresh package, which automatically generated hundreds of features from the time series data used in the feature extraction process, can be used to enhance the accuracy of the predictions.

In terms of effective features, the prediction model that considered only actual past yield data was the most accurate for both crops. The results of this experiment showed that past yields had the highest contribution to accurate predictions and

the environmental data were probably noise. We aim to further improve the accuracy of the models in future works by utilizing other prediction methods such as time series analysis.

## 5.2 Future Works

(1) Examination of feature quantities

As mentioned in the discussion, we are considering incorporating features related to human factors such as production management and shipment adjustment methods into the prediction model.

(2) Improvement of prediction accuracy by applying time series analysis method

Experiments showed that the contribution of past yields was extremely high, suggesting that time series analysis methods that specialize in predicting time series data, which can capture autocorrelation and cyclical variation, are more suitable. For the time series analysis method, the use of an auto regressive integrated moving average, which applies to nonstationary processes, will be considered.

## REFERENCES

[1] 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 10, 2020](*in Japanese*).

[2] Ministry of Agriculture Forestry and Fisheries, "Explanation of the Plant Factory, Japan Center for Social Development Research", http://www. maff. go. jp/j/heya/sodan/1308/01. html [Accessed December 10, 2020](*in Japanese*).

[3] Ministry of Agriculture Forestry and Fisheries, "Status of installation of horticultural facilities (2019)", https://www.maff.go.jp/j/seisan/ryutu/engei/sisetsu/haipura/setti_30.html[Accessed December 10, 2020](*in Japanese*).

[4] Ministry of Agriculture, Forestry and Fisheries, "Vegetable Production and Shipment Statistics(2018)", https://www.maff.go.jp/j/tokei/kikaku/book/seisan/attach/pdf/30_yasai- 1.pdf [Accessed December 10, 2020](*in Japanese*).

[5] Ministry of Agriculture Forestry and Fisheries, "Management Statistics by Farming Type(2016)", https://www.maff.go.jp/j/tokei/kikaku/book/seisan/attach/pdf/30_yasai- 1.pdf [Accessed December 10, 2020](*in Japanese*).

[6] WageningenUR, "Quantitative Information on Dutch grennhouse horticulture (2017)", https://www.wur.nl/en/newsarticle/Quantitative-Information-for-Greenhouse- Horticulture-KWIN-2016-2017.html[Accessed December 10, 2020](*in Japanese*).

[7] 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).

[8] A. Bashar and S. Pearson, G. Leontidis, and S.Kollias, "Using Deep Learning to Predict Plant Growth and Yield in Greenhouse Environments"(2019).

[9] W. C. Lin, and B. D. Hill, "Neural network modelling to predict weekly yields of sweet peppers in a commercial greenhouse", Canadian Journal of Plant Science, Vol. 88, pp. 531–536 (2008).

[10] M. Stas, J. Van Orshoven, Q. Dong, S. Heremans, and B. Zhang, "A comparison of machine learning algorithms for regional wheat yield prediction using NDVI time series of SPOT-VGT", IEEE, pp. 258–262 (2016).

[11] Pantazi, X., Moshou, D., Alexandridis, T., Whetton, R. and Mouazen, A., "Wheat yield prediction using machine learning and advanced sensing techniques", Computers and Electronics in Agriculture, Vol. 121, pp. 57 – 65 (2016).

[12] H. Li, Z. Chen, W. Wu, Z. Jiang, B. Liu, and T. Hasi, "Crop model data assimilation with particle filter for yield prediction using leaf area in- dex of different temporal scales", 2015 Fourth International Conference on Agro-Geoinformatics, pp. 401–406 (2015).

[13] G. Okuno, and 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*).

[14] W. C. Lin, D. Frey, G. D. Nigh, and C. Ying, "Combined Analysis to Characterize Yield Pattern of Greenhouse-grown Red Sweet Peppers", HortScience, Vol. 44, pp. 362–365 (2009).

[15] Y. Todate, M. Oba, and M.Takamori, "Prediction of weekly yields for fruit and vegetable crops in plant factories", Information Systems and Social Environment (IS) 149th Research and Presentation Meeting, Vol.151,No.12,pp.1-8 (2020).

[16] A. González-Sanchez, J. Frausto-Solis, and W. Ojeda, "Predictive ability of machine learning methods for massive crop yield prediction", SPANISH JOURNAL OF AGRICULTURAL RESEAR(2014).

[17] M. Christ, N. Braun, J. Neuffer, and A.W. Kempa-Liehr, "Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh -- A Python package)". Neurocomputing, Vol. 307, pp. 72-77(2018).

[18] M. Christ, A.W. Kempa-Liehr, and M. Feindt, "Distributed and parallel time series feature extraction for industrial big data applications", Asian Machine Learning Conference (ACML) 2016, Workshop on Learning on Big Data (WLBD), Hamilton (New Zealand)(2016).

**Yuki Todate** received his B.E. and M.E. degrees in infor-mation science from Future Univer-sity Hakodate, Japan in 2019 and 2021. Hers research interests include smart agriculture, plant factory, and statistical modeling. She currently works in Nomura Research Institute, Ltd..

**Michiko Oba** received her B.S. in physics from Japan Women's University in 1982, and Ph.D. in engineering from Osaka University, Japan, in 2001. She worked in the Systems Development Laboratory and the Software Division of Hitachi Ltd. Presently, she is a professor in the Department of Media Architecture, Future University Hakodate, Japan.Prof. Oba is a member of IEEE Computer Society, the Information Processing Society of Japan (IPSJ), the Institute of Electrical Engineers of Japan (IEEJ). She became a Council Member of Science Council of Japan in 2020. She received IPSJ fellow in 2020.

**Mitsuru Takamori** is president of IT consulting company called Agri. Connections Corp. from 2013 and he is doing research to utilize ICT and IOT in the field of agriculture.
1988-2013 Project manager in IBM Japan 2012-2016 Project Professor in Future University Hakodate.

<u>Regular Paper</u>

# A Personal Authentication Method Based on Eye Movement Trajectory

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. In order to overcome these weaknesses, this study focuses on the user's eye movement. As an example of existing research, Kinnunen et al. proposed an authentication method using features of unconscious eye movement when users are viewing a video [1]. In this method, it is difficult to update the registered authentication information. De Luca et al. proposed a password authentication method in which the user inputs a PIN by looking at number keys shown on a display [2]. This method has 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 conducted an experiment to evaluate features relating to fixation and saccade that are effective for classifying users. Fixation is the eye movement which people hold the line of sights, and saccade is the rapid eye movement to change the position of the sight. In this experiment, we used these features to classify users. The experimental result showed the proposed method improved classification accuracy when compared to our previous method. It was suggested that the fixation and saccade features were effective for user classification. Next, we conducted an experiment to examine a learning algorithm suitable for the proposed method. In this experiment, we used One Class SVM and Isolation Forest for personal identification. 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 for improvement of authentication accuracy. In addition, we need to investigate learning algorithms to improve identification accuracy.

*Keywords*: Personal authentication, Eye movement trajectory, Drawing feature, Classification of users, Error detection

## 1 INTRODUCTION

In recent years, mobile terminals such as laptops, smartphones, and tablets have become popular. Many users perform personal authentication on web services, applications, and online shopping. A variety of information is shared by mobile terminals. If the authentication information is leaked, there is a risk of it being used fraudulently. Therefore, it is important to improve the security level of authentication on mobile terminals.

Password authentication and biometric authentication are popular means of personal authentication, and are also used for authentication on mobile terminals. Password authentication is authentication that uses the user's knowledge such as passwords and PINs. Password authentication information cannot get lost because it does not require a physical object such as an IC card or key. However, authentication information can be leaked when someone looks over another person's shoulder in a public space. This is an act of information theft that is possible even if the attacker does not have specialized knowledge. Therefore, authentication information can be leaked to anyone. In addition, if the authentication information of the number of digits is small, 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. Physical features are information of body parts that are unique such as fingerprints and faces. Behavioral features are information of human behavior that can be reproduced by a person. Biometric authentication is robust against over-the-shoulder information theft and incurs less burden, in that users do not need to remember authentication information. 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 such as fingerprints and irises cannot be consciously altered. In authentication based on behavioral features, it is difficult to forge the authentication information because it does not use part of the body. Nevertheless, it is difficult to update authentication information consciously when unconscious habits are used as authentication information. If the authentication information is forged, it is difficult to deal with impersonation because users cannot update the authentication information. Therefore, biometric authentication has also difficulty dealing with impersonation if the biometric authentication information is forged once.

The weakness of password authentication is the ease of leakage of authentication information. In addition, the password authentication information be guessed easily. The weakness of biometric authentication is the difficulty of dealing with impersonation because it is difficult to update authentication information. These weaknesses must be overcome in order to perform secure authentication on mobile terminals. There are many works of research using physical

features 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], [2], [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. Also, user eye movement can be consciously reproduced. Authentication information can be updated by updating the user eye movement.

This study proposes a personal authentication method using a personal authentication based on eye movement trajectory, that is the trajectory drawn by 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, by using user eye movement. 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 [16]. 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 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 [3], [4]. Nakakuni et al propose a method that uses features of keystroke dynamics when users enter their surname for authentication [3]. In this method, it is thought that the input of a surname is highly reproducible and has a stable rhythm. Based on this hypothesis, the timing of the user's keystrokes is used for authentication. Zhou et al. used keystroke acoustic features in addition to keystroke dynamics features for authentication [4]. In this method, keystroke acoustics are collected with a microphone. In addition, they calculate MFCC (Mel-Frequency Cepstral Coefficients) by the acoustics and used for authentication.

There are studies using walking features for authentication [5], [6]. Li et al. use walking features while holding a mobile phone for authentication [5]. This method extracts features of walking by an accelerometer which is mounted on the mobile phone and calculates statistical features for au-

thentication. Musale et al. extract leg and arm movement features during walking and use these features for authentication [6]. This method uses a smartwatch or smartphone to extract features related to human behavior. These features make it possible to classify users by a small number of features.

There are studies using the features of smartphone operation for authentication [7], [8]. Salem et al. use keystrokes as a second authentication factor when performing authentication with a touchscreen terminal [7]. In this method, they have developed a virtual keyboard and 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 [8]. Features such as flicking and shaking of the terminal during text input are used for continuous authentication.

These methods use unconscious habits and patterns that appear in each user's behavior as authentication information. Therefore, we think that it is difficult to update such authentication information and deal with impersonation.

### 2.2 Authentication Based on Eye Movement Features

There are studies using unconscious or conscious eye movement for authentication [1], [2], [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. have proposed an authentication method using eye movement and head movement [9]. In this method, authentication is performed by displaying random visual stimuli and measuring the unconscious eye movements and head movements with a camera.

These studies do not make users aware of authentication during authentication. However, in these methods, it will be difficult to update the registered authentication information. Therefore, we think that it is difficult to deal with impersonation.

As studies using conscious eye movement for authentication, there are studies performing authentication using password by eye movement [2], [10]. In addition, there is a study that performs authentication by having the user draw with their eye movement trajectory [11]. De Luca et al. proposed a password authentication method by gazing at PIN keys shown on a display [2]. In this method, the user enters a PIN code by gazing at numbers on the keypad shown on the display and authentication is performed. Khamis et al. have proposed a personal authentication method that combines passwords and eye movement information [10]. This method uses multimodal passwords with touch input and gaze direction (e.g., left-3-right-4) for authentication. 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]. In this method, users select one of the characters from a set of characters to be used as the authentication information. This study uses characters and symbols that can be registered for authentication information. The features such as drawing time and drawing speed extracted from these characters are used for authenti-

cation. It is possible to update the authentication information by updating the characters registered as authentication information.

In these methods, it is possible to update authentication information and deal with impersonation. However, we think that in the methods used in [2] and [10], authentication information is simple if the number of authentication password digits is small. In the method used in [11], the characters that can be used as authentication information are limited. The methods in [10] and [11] use alphabets and numbers that are well known for everyone as authentication information. These authentication information can be guessed by attackers because alphabets and numbers are usually used by many users on a daily basis and are familiar for them. On the other hand, the proposed method uses the user's own defined eye movement trajectory as authentication information. It is difficult for an attacker to guess the authentication information of the proposed method because each user defines the trajectory information independently and the authentication information is not formed from well-known information such as alphabets and numbers. Accordingly, we think the proposed method are more robust for attackers compared to the methods in [10] and [11].

## 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. Our method prevents information theft or forgery because eye movement information is not visible. In addition, authentication information can be updated by updating eye movement, because the user's eye movement is consciously reproduced. Therefore, we think it is possible to deal with impersonation.

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 because authentication information does not depend on alphabets and numbers. If only the shape of the eye movement trajectory is used for authentication, there is a risk of impersonation because the shape of trajectory does not contain personal features. We introduce the authentication based on drawing features to makes the proposed method more robust against impersonation.

The goal of this paper is to investigate the combination of features related on local eye movement that are effective for the proposed method, and investigate learning algorithms suitable for 1:1 authentication.

### 3.2  Proposed System

In this section, we explain the structure of the proposed system. Fig. 1 shows an overview of the proposed system.
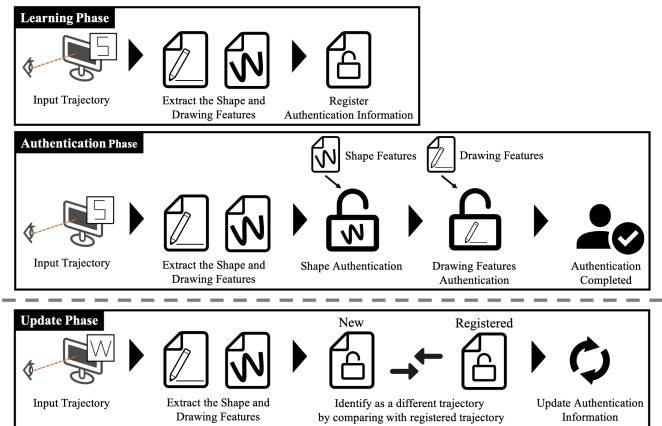


Figure 1: An overview of the proposed system

The proposed system consists of a learning phase, and an authentication phase and an update phase. Features are extracted from the eye movement trajectory and these features are used in each phase. The learning phase is the phase of creating the authentication information with multiple entries. Nothing is displayed on the screen during drawing the eye movement trajectory. During drawing the eye movement trajectory, neither authentication information nor guides for drawing the trajectory are displayed. 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.

The authentication phase is the phase which authentication is performed. In this phase, users enter the eye movement trajectory registered on the learning phase on a blank screen for authentication. First, users perform the authentication based on the shape of the user's eye movement trajectory. If the authentication is successful, the authentication based on drawing features is performed as the next step. The authentication of the proposed method is completed by success of these two steps authentication.

The update phase is the phase which authentication information is updated. In the phase, users input eye movement trajectory which users want to newly register. After extracting features, the newly entered authentication information is compared with the registered authentication information. If the new trajectory is identified as a different trajectory, authentication information is updated.

### 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 eyeline measurement devices, there are contact type devices such a glasses-shaped

device, and contactless devices such as a desktop device [17]. 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. In this study, we assume that the proposed method will use a camera on a mobile terminal to perform eye tracking in the future. If users use contactless type devices, the burden on the user during authentication is low because users are not required to wear the devices. Therefore, the proposed method uses a contactless type device.

**Approach to research task 2**

We use features that can be extracted from the coordinates data of the eye movement trajectory. Since the proposed system uses the eye movement trajectory defined by the user, it is necessary to estimate the shape of trajectory. In our previous study [16], we classified trajectories using each feature that can be extracted from coordinates data and images in order to investigate features that are effective for estimating the shape of trajectory. 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 for estimating the shape of trajectory.

**Approach to research task 3**

We use fixation and saccade features during eye movement for classifying users. We classify users in order to identify the registered users and protect against impersonation. 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. Fixation is the eye movement that make the line of sight move to fixing in position. Saccade is the rapid eye movement that change the points of sight. We think we can extract features which are more effective for classifying users, by detecting features of local eye movement.

**Approach to research task 4**

We use an error detection algorithm. Error detection is the method that learns only normal data and identifies whether unknown data is normal data or error data. There are 1:1 authentication and 1:N authentication, as authentication methods. 1:1 authentication uses only the user's data for learning and identifies whether the person attempting to access the device is the original user or attackers. 1:N authentication identifies the as one user from among all the registered users. The proposed method performs 1:1 authentication because we assume that the proposed method is applied to mobile terminals retained by people. Therefore, we think that the error detection algorithm is effective for the user identification in 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 (Fig. 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. draw an eye movement trajectory not to move their heads. 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 Fig. 3. An example of the collected data is shown in Fog. 4.

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.
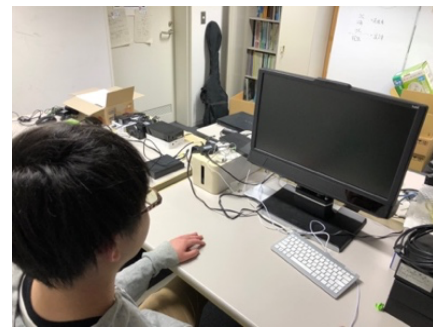


Figure 2: Measurement equipment (Tobii Pro Tx-300)



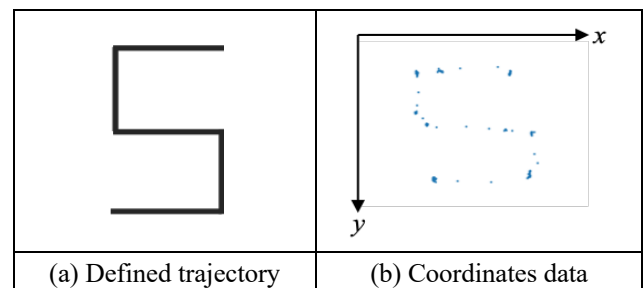Figure 3: Measurement environment



| (a) Defined trajectory | (b) Coordinates data |
| --- | --- |

Figure 4: An example of collected 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. However, large shakes may become outliers. Therefore, we perform preprocessing before the authentication based on shape and drawing features.

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. Shakes of eye movement and fixations are removed by this process. A smaller number of divisions can remove shakes of eye movement and maintain the approximate shape information of trajectory. A larger number of divisions can remove large outliers while maintaining the original shape information of trajectory. Therefore, the smaller number of divisions will be suitable for the authentication based on shape. The larger number of divisions will be suitable for the authentication based on drawing features. We extract features from the average coordinates data calculated with each division.

## 3.6　Investigation of the Features for the Proposed Method

In this section, we explain drawing features using for personal identification.

### 3.6.1　Investigation of Drawing Features for Personal Identification

In this paper, we use fixation and saccade features for personal identification. Fixation is an eye movement that is performed to fix the direction of sight. Saccade is a rapid eye movement that abruptly changes the point of fixation [12], [13]. A representative example of saccade is the eye movement during line transitions while reading. If we can capture individual differences in these eye movements, we will be able to use these differences as personal features for authentication. The researches [14], [15] have discussed the individual differences in eye movements by saccade. The proposed method extracts these eye movement features to use for personal identification. 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

In the proposed method, we extract fixation and saccade features from average coordinates data.

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 draw-
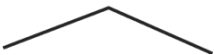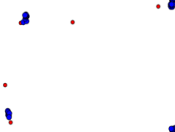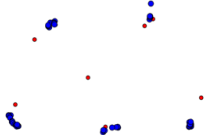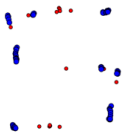
ing trajectory by the eye movement. Therefore, we calculate the features from all fixations detected.

Second, we extract saccade features. The trajectory has a beginning and an end point, with a turning point in the middle. Since the shape is drawn by connecting these points, 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.

Blue points are where fixations were observed. Blue points are crowded at the starting and, finishing point of the trajectory and turning points. The result shows that the proposed method can detect fixation points.

Table 1: Fixations extracted from trajectory

| Trajectory | Fixations |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |

### 3.6.3    Features Used for Classifying Users

Table 2 shows features used for the proposed method and our previous study [16]. Previous study used the global features such as the amount of change in the average coordinates data. The proposed method does not use the global features, but fixation and saccade features are used as local features. We think that the amount of change in the average coordinates data has redundant information for classifying users because its data contains the data of all frames when users draw an eye movement trajectory. Therefore, we think that we can extract eye movement features which do not contain redundant information by using fixation and saccade 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.

## 3.7    Investigation of Learning Algorithm

The proposed method uses an error detection algorithm for building a learning model. There are studies using error detection algorithms for personal authentication [8], [18]. 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. We treat users as normal data and attackers as error data.

Table 2: Features used for experiments

| Features | Previous study [16] | 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 | × | ○ |

## 4    EVALUATION

We explain the experiment using local features of eye movement for classifying users in Section 4.1. In addition, we explain the experiment using error detection algorithms for personal identification in Section 4.2.

## 4.1    Classifying Users Using Local Features

In the proposed method, authentication information is created for each individual. The features that can capture the characteristic differences of the individual are desirable for personal authentication. The individual differences are effective not only personal authentication but also personal classification. In other words, if we cannot find the individual differences that are effective features for the classification, it will be difficult to realize personal identification. Therefore, we conducted the experiment about classification of five subjects by using Random Forest to examine the features effective for the classification. In this classification, we used fixation and saccade features. We instructed the subjects to draw the trajectory shown in Fig. 5 30 times. Table 2 shows the features used for classifying users in the proposed method and our previous study [16]. We classify users and calculate the F-measure to evaluate classification accuracy by 10-fold cross-validation. The training and test data in cross-validation include data for all subjects. In addition, 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. 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. The proposed method improved the classification accuracy. This method does not use the amount of change in average coordinates data, and adds the fixation and saccade features for classifying users. Therefore, we think that these features improve the classification accuracy.
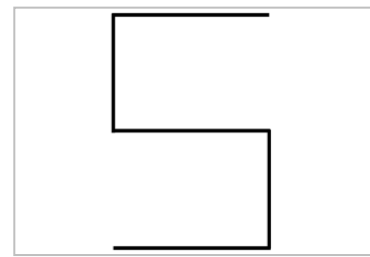


Figure 5: Trajectory used for the experiments

Table 3: Environment of the experiments

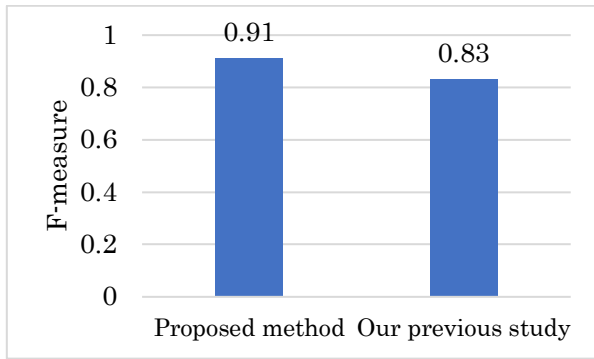| Items | Specification |
|---|---|
| CPU | Intel Core i5 2.4GHz |
| OS | macOS Catalina10.15.2 |
| Language | Python3.4.5 |
| Library | scikit-learn0.18.1 |

Figure 6: Result of classifying users

Table 4 shows the top 10 features of the proposed method with variable importance. The variable importance with related on saccade is high. This shows 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. The variable importance of variance and standard deviation of x-coordinates are high in the proposed method. The proposed method does not use the amount of change in the coordinates for the classifying users. Therefore, we think that other features contribute to classification.

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. We think it is necessary to extract and add other local eye movement features to improve classification accuracy.

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 | - |

## 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 Fig. 5 30 times. We use the features shown in Table 2 for this experiment. We evaluate the identification accuracy by F-measure, Precision, Recall, Specificity, FAR (False Acceptance Rate), and FRR (False Rejection Rate). Precision is the percentage of data that predict the user's identity among the data that predict the user. Recall is the percentage of data that are predicted to be the user among the data that are actually the user. Specificity is the percentage of data that are predicted to be others among the data that are actually others. FAR is the probability of mistakenly identifying others (non-users) as authentication users. FRR is the probability of mistakenly identifying users as others. The process for calculating the identification accuracy is as follows:

(1) We perform a test to identify the attacker, using one subject's data as the training data and the other subject's data as the evaluation data to calculate the accuracy. The same process is applied to each subject, and the average accuracy of the five subjects is calculated. This test calculates the number of correctly rejected attackers (true negative) and the number of incorrectly accepted attackers (false positive). The Specificity and FAR are calculated based on these calculated results.

(2) We perform a test to identify the user, and the accuracy is calculated by performing a 10-fold cross-validation using the data of a single subject. The same process is applied to each subject, and the average accuracy of the five subjects is calculated. This test calculates the number of correct acceptances (true positive) and false rejections (false negative). The Recall and FRR are calculated based on these results.

We calculate the F-measure and Precision the accuracy of the personal identification. The accuracy calculated in this way is summarized in Table 5.

This result showed that the personal user was accepted with relatively high accuracy in the case of Isolation Forest. But, the identification accuracy does not reach the level enough to perform authentication. We think that we were unable to build a learning model which is effective for personal identification, because there was not sufficient learning data. Therefore, solving the lack of learning data is one of our future tasks. The results of the evaluation suggested that Isolation Forest is effective as an algorithm for personal identification in our study.

Table 5: Result of personal identification

| | One Class SVM | Isolation Forest |
| --- | --- | --- |
| F-measure | 0.61 | 0.73 |
| Precision | 0.86 | 0.75 |
| Recall | 0.49 | 0.75 |
| Specificity | 0.97 | 0.91 |
| FAR | 0.03 | 0.08 |
| FRR | 0.51 | 0.27 |

We performed feature selection using the stepwise method to evaluate which features were effective for personal identification using Isolation Forest. We adopted the stepwise method that is one of representative feature selection methods. The feature selection method searches for the best combination of features by adding or deleting features one by one. We used the selecting and forward feature selection method. The method is one of the stepwise methods. This method increases features one by one from the initial state with no features. We can see which features contribute it by checking the order of addition of the features based on this method.

The process for feature selection in this method is as follows:
(1) Add one feature from among the unselected features, perform personal identification and calculate the F-measure.
(2) After calculating the F-measure, undo the added features and add another feature.
(3) Perform the steps (1) and (2) using all the features, and actually add the features that have the highest F-measure.
(4) Return to the step (1) and repeat the steps (1) ~ (3) until all the features have been added.

The process can be used to evaluate which features are effective for personal identification. The result is shown in Fig. 7.
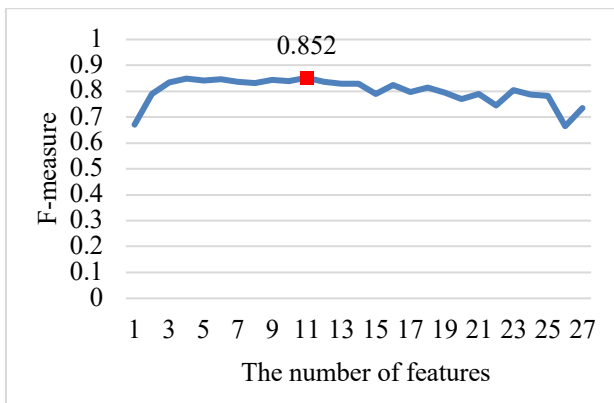


Figure 7: The change of F-measure over the number of features

Table 6: List of features when F-measure is the highest

| Order of addition | Features |
|---|---|
| 1 | Standard deviation of x coordinates |
| 2 | Drawing time |
| 3 | Standard deviation of y coordinates |
| 4 | Average speed of saccade in y direction |
| 5 | Number of occurrences of fixation |
| 6 | Minimum speed of saccade in x direction |
| 7 | Minimum variance of fixation in x direction |
| 8 | Maximum speed of saccade in x direction |
| 9 | Minimum standard deviation of fixation in x direction |
| 10 | Variance of y coordinates |
| 11 | Variance of x coordinates |

When the number of the added features was 11, the F-measure was the highest and the value was 0.852. At this time, the FAR was 0.008 and the FRR was 0.23. The features at this time and the order of addition are shown in Table 6. This table shows that these features are effective for personal identification. The fixation and saccade features are included in these effective features. Comparing Table 4 with Table 6, the features that occur in both tables are Average speed of saccade in y direction, Minimum speed of saccade in x direction, Maximum speed of saccade in x direction. These features showed significant individual differences. This experimental result suggested that these features were effective in authentication based on drawing features. The number of occurrences of fixation was included only in Table 6 as a fixation feature. In this experiment, there were few individual differences in the fixation features. In the trajectory used for the experiment, individual differences appeared in saccades. However, individual differences did not appear in fixation because the subjects did not perform fixation much and drew smoothly. In addition, the number of fixation and saccade features that were effective in this study was 6 out of 22 dimensions. In order to increase the accuracy of the personal identification, it is necessary to re-examine and add the features that are effective for personal identification.

## 4.3　Discussions and Future Works

The goal of this research is to realize an authentication for mobile terminals such as laptop PCs and smartphones. Applications that we assume include unlocking of mobile terminals themselves and login to services from mobile terminals. In public spaces such as stations and cafes, it is important to prevent unauthorized login by others while the user is away from the terminal, or login by others when the terminal is lost. In the experiment in Section 4.2, we performed personal identification and feature selection. As a result, in personal identification using the features with the highest accuracy, the F-measure was 0.852, FAR was 0.008 and FRR was 0.23. Related researches [3], [7], [9], for similar applications (to mobile terminals) have achieved FAR of less than 0.1. We think that regarding FAR, we have achieved a satisfactory level in comparison with these researches. On the other hand, the FRR in the related researches [3], [7], [9], is less than 0.1, but the FRR in the proposed method is 0.23. It is difficult to say that the accuracy is sufficient at this point. The reason for the low accuracy may be that each subject cannot reproduce the same eye movement trajectory. As a countermeasure, we are considering improvement of reproducibility of the eye movement trajectory by displaying a reference point during the drawing process.

## 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 sac-

cade 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. We compared the classification accuracy of the proposed method with the classification accuracy of our previous study. 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, Precision, Recall, Specificity, FAR, and FRR. In the case of using One Class SVM, the F-measure was 0.61, Precision was 0.86, Specificity was 0.97, Recall was 0.49, FAR was 0.03 and FRR was 0.51. In the case of using Isolation Forest, the F-measure was 0.73, Precision was 0.75, Recall was 0.75, Specificity was 0.91, FAR was 0.08 and FRR was 0.27. 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.

There are three future tasks for this study. First, we need to consider other features about local eye movement and add these to improve classification accuracy. Next, we will consider a method to solve the problem of lack of learning data for improvement of personal identification accuracy. Finally, we will further investigate learning algorithms to improve the accuracy of each authentication.

## ACKNOWLEDGEMENTS

## REFERENCES

[1] T. Kinnunenn, 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] A. De Luca, R. Weiss and H. Drewes, "Evaluation of Eye-Gaze Interaction Methods for Security Enhanced PIN-Entry," Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, OZCHI'7, pp.199-202 (2007).

[3] 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).

[4] Q. Zhou, Y. Yang, F. Hong, Y. Feng and Z. Guo, "User Identification and Authentication Using Keystroke Dynamics with Acoustic Signal," 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), pp.445-449 (2016).

[5] H. Li, J. Yu and Q. Cao, "Intelligent Walk Authentication: Implicit Authentication When You Walk with Smartphone," IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp.1113-1116 (2018).

[6] 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).

[7] 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).

[8] S. Itou 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).

[9] 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).

[10] 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 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp.2156-2164 (2016).

[11] H. Mukai and T. Ogawa, "Feature Extraction of Eyegaze Path for Personal Authentication," IPSJ Transaction on Digital Contents (DCON), Vol.4, No.2, pp.27-32 (2018) (in Japanese).

[12] K. Ukai, "Eye Movement: Characteristics and Method of Measurement," Japanese Journal of Optics, Vol.23, No.1, pp.2-8 (1994) (in Japanese).

[13] Q. Yang, P. M. Bucci and Z. Kapoula, "The Latency of Saccades, Vergence, and Combined Eye Movements in Children and in Adults," Investigative Ophthalmology & Visual Science, Vol.43, No.9, pp.2939-2949 (2002).

[14] G. Bargary, J. M. Bosten, P. T. Goodbourn, A. J. Lawrance-Owen, R. E. Hogg and J. D. Mollon, "Individual Differences in Human Eye Movements: An Oculomotor Signature?," Vision Research, Vol.141, pp.157-169 (2017).

[15] B. de Haas, A. L. lakovidis, D. S. Schwarzkopf and K. R. Gegenfurtner, "Individual Differences in Visual Salience Vary Along Semantic Dimensions," Proceedings of the National Academy of Sciences," Vol.116, No.24, pp.11687-11692 (2019).

[16] T. Fujimoto and Y. Shiraishi, "An Examination of Personal Authentication Method Using Shape of Gaze Trajectory and Drawing Features," Proceedings of Multi-

media, Distributed, Collaborated and Mobile Symposium of IPSJ, pp.1423-1432 (2019) (*in Japanese*).

[17] Tobii Pro Mechanism of Eye Tracker, tobii pro, https://www.tobiipro.com/ja/service-support/learning-center/eye-tracking-essentials/how-do-tobii-eye-trackers-work/ (accessed 2019-05-06).

[18] 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," Advances on Broad-Band Wireless Computing, Communication and Applications, Lecture Notes in Networks and Systems (LNNS), Vol. 97, Springer, pp.352-362 (2019).

**Takumi Fujimoto** received his B.E. and M.E. degrees in information science from Future University Hakodate, Japan in 2019 and 2021. His research interests include mobile sensing, security and machine learning. He



**Yoh Shiraishi** received doctor's degree from Keio University in 2004. He is currently a professor at the Department of Media Architecture, School of Systems Information Science, Future University Hakodate Japan. His re-

## Submission Guidance

### About IJIS

International Journal of Informatics Society (ISSN 1883-4566) is published in one volume of three issues a year. One should be a member of Informatics Society for the submission of the article at least. A submission article is reviewed at least two reviewer. The online version of the journal is available at the following site: http://www.infsoc.org.

### Aims and Scope of Informatics Society

The evolution of informatics heralds a new information society. It provides more convenience to our life. Informatics and technologies have been integrated by various fields. For example, mathematics, linguistics, logics, engineering, and new fields will join it. Especially, we are continuing to maintain an awareness of informatics and communication convergence. Informatics Society is the organization that tries to develop informatics and technologies with this convergence. International Journal of Informatics Society (IJIS) is the journal of Informatics Society.

Areas of interest include, but are not limited to:

| | |
|---|---|
| Internet of Things (IoT) | Intelligent Transportation System |
| Smart Cities, Communities, and Spaces | Distributed Computing |
| Big Data, Artificial Intelligence, and Data Science | Multi-media communication |
| Network Systems and Protocols | Information systems |
| Computer Supported Cooperative Work and Groupware | Mobile computing |
| Security and Privacy in Information Systems | Ubiquitous computing |

### Instruction to Authors

For detailed instructions please refer to the Authors Corner on our Web site, http://www.infsoc.org/.

Submission of manuscripts: There is no limitation of page count as full papers, each of which will be subject to a full review process. An electronic, PDF-based submission of papers is mandatory. Download and use the LaTeX2e or Microsoft Word sample IJIS formats.

http://www.infsoc.org/IJIS-Format.pdf

LaTeX2e

LaTeX2e files (ZIP) http://www.infsoc.org/template_IJIS.zip

Microsoft Word<sup>TM</sup>

Sample document   http://www.infsoc.org/sample_IJIS.doc

Please send the PDF file of your paper to secretariat@infsoc.org with the following information:

Title, Author: Name (Affiliation), Name (Affiliation), Corresponding Author. Address, Tel, Fax, E-mail:

### Copyright

For all copying, reprint, or republication permission, write to: Copyrights and Permissions Department, Informatics Society, secretariat@infsoc.org.

### Publisher

Address:   Informatics Laboratory, 3-41 Tsujimachi, Kitaku, Nagoya 462-0032, Japan

E-mail:   secretariat@infsoc.org

# CONTENTS